

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 5, 2006, 23:14:07 ; Search time 437 Seconds  
(without alignments)  
6551.022 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 5349  
Sequence: 1 MSSNPFLEKGFVHSLDPE.....RHMEHFLQTNVYKPGMSFA 1020

## Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/abes/ABSSWEB\_spool/US10645746/runat\_05072006\_110149\_5129/app\_query.fasta\_1  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs08 -USER=US10645746 @CGN 1 1 271 @runat\_05072006\_110149\_5129 -NCFU=6  
-ICFU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued Patents NA:\*  
1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1\_COMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5\_COMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A\_COMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B\_COMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7\_COMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H\_COMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PCTUS\_COMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP\_COMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/RE\_COMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	577	10.8	7478	3	US-09-949-016-948 Sequence 948, App
2	577	10.8	7478	3	US-09-949-016-948 Sequence 4453, App
3	575.5	10.8	2910	3	US-09-533-029-105 Sequence 105, App
4	452.5	8.5	1662	3	US-09-270-767-11928 Sequence 11928, A
5	388	7.3	1874	3	US-09-270-767-11928 Sequence 12369, A
6	312	5.8	2328	3	US-10-043-774B-1 Sequence 1, Appli
7	306	5.7	3472	3	US-09-873-737A-5 Sequence 5, Appli
8	304	5.7	4064	3	US-09-873-737A-3 Sequence 3, Appli

9	292	5.5	3620	4	US-10-094-749-902 Sequence 902, App
10	288	5.4	2579	3	US-10-104-047-1469 Sequence 1469, Ap
11	278.5	5.2	3138	3	US-10-104-047-1676 Sequence 1676, Ap
12	265	5.0	771	3	US-09-248-796A-6819 Sequence 6819, Ap
13	256.5	4.8	734	4	US-09-297-648-2528 Sequence 2528, Ap
14	205	3.8	3047	3	US-09-873-737A-1 Sequence 1, Appli
15	178.5	3.3	2167	3	US-09-270-767-13282 Sequence 13282, A
16	170	3.2	300	4	US-09-297-648-1917 Sequence 1917, Ap
17	167	3.1	9626	3	US-09-150-867-2 Sequence 2, Appli
18	167	3.1	9626	5	US-09-724-584-2 Sequence 1, Appli
19	165.5	3.1	1664976	3	US-08-916-421B-1 Sequence 1, Appli
20	165.5	3.1	1664976	3	US-09-692-570-1 Sequence 130, App
21	164.5	3.1	8503	3	US-09-620-312D-130 Sequence 2009, Ap
22	163	3.0	6276	3	US-09-949-016-2009 Sequence 1028, Ap
23	163	3.0	6284	3	US-09-949-016-1028 Sequence 5831, Ap
24	162.5	3.0	5192	3	US-09-949-016-5831 Sequence 27, Appl
25	162.5	3.0	6773	3	US-09-166-350-27 Sequence 30, Appl
26	161.5	3.0	8257	3	US-09-595-684B-30 Sequence 5562, Ap
27	158	3.0	8530	3	US-09-949-016-5562 Sequence 636, App
28	158	3.0	10300	3	US-09-949-016-636 Sequence 17304, A
29	158	3.0	38575	3	US-09-949-016-17304 Sequence 12378, A
30	158	3.0	119153	3	US-09-949-016-12378 Sequence 2, Appli
31	157	2.9	10136	2	US-08-353-700-2 Sequence 12, Appl
32	157	2.9	10136	7	PCT-US95-16216-2 Sequence 12, Appl
33	156	2.9	4868	2	US-08-139-937-12 Sequence 12, Appl
34	156	2.9	4868	7	PCT-US93-11310-12 Sequence 5, Appli
35	156	2.9	8789	2	US-08-328-254-5 Sequence 117, App
36	155	2.9	6921	3	US-09-643-597-117 Sequence 117, App
37	155	2.9	6921	3	US-09-480-884A-117 Sequence 117, App
38	155	2.9	6921	3	US-09-542-615A-117 Sequence 117, App
39	155	2.9	6921	3	US-09-506-421B-117 Sequence 117, App
40	155	2.9	6921	3	US-09-221-107-117 Sequence 117, App
41	155	2.9	6921	3	US-09-466-396A-117 Sequence 117, App
42	155	2.9	6921	3	US-09-476-496A-117 Sequence 117, App
43	155	2.9	6921	3	US-09-630-940B-117 Sequence 117, App
44	155	2.9	6921	3	US-09-285-479-117 Sequence 117, App
45	155	2.9	6921	3	US-10-007-700-117 Sequence 117, App

## ALIGNMENTS

RESULT 1  
US-09-949-016-948  
; Sequence 948, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 948  
; LENGTH: 7478  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-948

Alignment Scores: 1.17e-55 Length: 7478  
Pred. No.: 577.00 Matches: 237  
Score: 40.2% Conservative: 173  
Percent Similarity: 23.2% Mismatches: 370  
Best Local Similarity: 10.8% Indels: 243  
Query Match: 3 Gaps: 38  
DB:

US-10-645-746-3 (1-1020) x US-09-949-016-948 (1-7478)

Qy	21	MetLysTrpLeuAlaAArgProThrGlyLysCysAspGlyLysPheTyrGlu-----	37
Db	209	ATGGATGGAA- GCGGGACCTCGGGAGCAGCTCGGGCGCTTACTCTGCCCCCTCGAC	267
Qy	38	-----LysLysValLeuLeu 42	
Db	268	CAGGTGTTCCAGGCACCTCGCGCGCTGGCAATTGGCACTGTGGGAAACCAATCAAGCTC	327
Qy	43	LeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGluTyrGlu	62
Db	328	CTGGCCAAATTACTTTGAGGTGGACATCCCTTAAGATCGAGCTGTACCACTACGAGGTGCAC	387
Qy	63	VallysMetThrLysGluValLeuAsnArgLysProGlyLysPheProLysLysThr 82	
Db	388	ATCAAGCCGGATAAG-----TGTCCTCCGCTAGAGTC 417	
Qy	83	GluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuAArgHisGluLysLys 102	
Db	418	AACCGGAGTGTGGAA-----TACATGCTCCAGCATTTCAAGCCT 458	
Qy	103	GlnThrAspPheIleLeuGluAspTyr-----ValPheAspGluLysAspThrValTyr 120	
Db	460	CAG-----ATCTTTGGTGATCGCAAGCCTGTGTATGATGAAAGAAACAATTATC 510	
Qy	121	SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysVal----- 138	
Db	511	-----ACTGTCAACAGCATGCCCCATTGGCAACGAGCGGTTCGACTTT 552	
Qy	139	---VallysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeuTyr 157	
Db	553	GAGGTGACAACTCTTGGGAAGGAGGAT---CGAATCTTAAAGTCTCCATCAAGTGG 609	
Qy	158	ThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerAArgLysAsnProGlu 177	
Db	610	CTAGCCATTGTGAGCTGGCGAAATGCTGCATCAGGCCCTGGTGCAGCGCCAGATCCCTGTT 669	
Qy	178	LysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysVal 197	
Db	670	CCCTTGGAGTCTGTCAAGCCCTCGATGTGCCCATGAGCCACCTGGCATCC-----ATG 723	
Qy	198	ArgTyrAlaProPheValAsnGluLulileLysValGlnPheAlaLysAsnPheValTyr 217	
Db	724	AGGTACACCCCTGTG----- 738	
Qy	218	AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnAArgPheGluGln 237	
Db	739	---GGCCGCTCCTTCTCTCACCGCTGAGGGCTACTACCACCG----- 780	
Qy	238	SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyLysGlu 257	
Db	781	-----CTGGGGGTGGCGGAGATCTGGTTCGGCTTTCCACAGTCTGTGGCGCCT 831	
Qy	258	LeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAla 277	
Db	832	CCCATG---TGGAGATGATCTCAACATTCATGCTCTCAGCCACTGCGCTTTTATAGGCA 888	
Qy	278	ProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSerCysAsnAsp 297	
Db	889	-----CAGCCAGTGATTGAGTTCATGTGTGAGGTGTGGACATCAGAACATAGATGAC 942	
Qy	298	AspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAla 317	
Db	943	CAGCCCCAAG-----CCCTCACCGACTCT 966	
Qy	318	AlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLys-----CysAla 334	
Db	967	CAGCGCGTTCCTTCCACCAAGAGATCAAGGCCCTGAAGGTGGAGATCACCCCTGTGGA 1026	
Qy	335	GluValTrpAspAsn-----GluMetSerArgLeuThrGluAArgHisLeuThr 350	

1027	Db	CAGATGAAGAGGAGATGCCCGGTGTGTAATGTTACCGGTGCCCTGCTACGCCATCAGACA	108
351	Qy	PheLeuAspLeuCysGluGluAenSerLeuValTyrLysValThrGlyLysSerAspArg	370
1087	Db	TTCCCTTTACAGCTGGAGAGTGGACAGCTGTG	1119
371	Qy	GlyArgAsnAlaLysLysYrAspThrLeuPheLysIleTyrGluGluAenLysLys	390
1120	Db	-----GAGTGACAGTGGCACAGTATTCAAGCAGAGAATAATAAC	1158
391	Qy	Phe---IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTyr	409
1159	Db	CTTCAGCTCAAGTATCCCATCTGCCCTGCTACAGTTGGCCAGGAACAAAGCATACC	1218
410	Qy	AlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArgIle	429
1219	Db	TACCTTCCCTAGAGTCTGTAACATT--GGCTGGGCGAGCGCTGTATTAAAGAACGTG	1275
430	Qy	AspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLys	449
1276	Db	ACGACAACACAGACCTCGACCATGATTAAGGCCACAGCTAGATCCGCTCCAGACAGACAG	1335
450	Qy	GluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAenPheVal	469
1336	Db	GAGGAGATCAGTCGCTGATGAAGAATGCCAGCTACAACCTTAGATCCC-----TACATC	1389
470	Qy	GluArgPheGlyLeuCysSerLysLeuGlnMetIleGlyCysProGlyLysValLeuLys	489
1390	Db	CAGGAATTTGGGATCAAAGTGAAGATGACATGACCGAGGTGACAGCGCGAGTCTGCCG	1449
490	Qy	GluProMetLeuValAenSerValAenGluGlnIleLysMetThrPro-----	505
1450	Db	CGGCCATCTTGCAGTACGCGCGCGCGAACC GGCGCCATTGCCACACCCCAATCAGG GTGC	1509
506	Qy	ValIleArgGlyPheGlnGluLysGlnLeu---AsnValValProGluLysGluLeu	523
1510	Db	TGGGACATGCGGGG-----AAACAGTTCTACAATGGGATTCAGATCAAGTCTGG	1560
524	Qy	CysCysAlaValPheValAenGluThrAlaGlyAenProCysValLeuGluAenAsp	543
1561	Db	GCCATCGCTGCTTCGCACCCCAAAA-----CAGTGTGGAGAGAG-----	1602
544	Qy	ValValLysPheTyrThrGluLeuIle-Gly-----GlyCysLysPhe	557
1603	Db	GTGCTCAAGAACTTCACAGACCACTGCGGAGAGATTTCAAGAGTGGCGGATGCGCTATC	1662
557	Qy	eArgGlyLeaArgIleGlyAlaAenGluAenArgGlyAlaGlnSerIleMetTyrAspAl	577
1663	Db	CAGGTCACACTGTTTCTGCAATATGCAAGGGGCA-GACAGCGTG-----	1710
577	Qy	aThrLysAsnGluTyrAlaPheTyrLysAenCysThrLeuAenThrGlyIleGlyArgPhe	597
1710	Db	-----	1710
597	Qy	eGluIleAlaAlaThrCluAlaLysAenMetPheGluArgLeuProAspLysGluGlnLys	617
1711	Db	-----GAGCTATGTTCCCGCATCTTCAGAGACACCTACTCAGG	1748
617	Qy	sValLeuMetPheIleIleSerLysArgGlnLeuAenAlaTyrGlyPheValLysHi	637
1749	Db	GCTGCAGCTCATATTGTCATCTGCCAGGGAAGACCGCGGTGTGCTGAGGTGAAACG	1808
637	Qy	sTyrCysAspHisThrIleGlyValAlaAenGlnHisIleThrSerGluThrValThrLys	657
1809	Db	TGTCGAGATACACTCTTGGGAATGGCTACGCACTGTGTGCAGTGAAGAACCTGTGCTCAA	1868
657	Qy	sAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLe	677
1869	Db	GACC-----TCACTCAGACTCTGTCCCAACCTCTGCCT	1901
677	Qy	uLysIleAsnAlaLysLeuGlyGlyIleAenGlnLeuAspTrpSerGluIleAlaGl	697
1902	Db	CAGATCAATGTCAACTTGGTGGCATTAACAATCATCTTA-----	1941



```
QY 697 uileSerProGluGluLysGluArgArgLysThrMetProLeuThrMetTyrValGlyI 717
Db 1942 ----GTCCACACAGCGCTCTGCGCTTTTCAACAGCCAGTG---ATATTCTGGGAGC 1994
QY 717 eAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaValValAl 737
Db 1995 AGATGTACACACCCACAGCGGATGGGAAACCTTCTATCACAGCAGTGGTAGG 2054
QY 737 aSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluCysAr 757
Db 2055 CAGTATGGATGCCACCCAGCCGATCTGTGCTACTGTGGGTACAG-----2103
QY 757 gProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGlu-----Al 775
Db 2104 -----CGACCCACGCAAGAGATCATTTGAAGACTTGTGTC 2135
QY 775 aLysPheValLys---LeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAl 794
Db 2136 CTACATGGTGGTGGTCTCTCATCCAAATCTACAGTCCACCGGTTTCAG---CCTAC 2192
QY 794 aHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAs 814
Db 2193 CCGCATCATCTCTACGAGATGGGTGCTGAAGCCAGCTACCCAGATACCTCCATT 2252
QY 814 pGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAs 834
Db 2253 TGAGCTACTGGCCATTTGCTGATCGCTCATCAACTG-----GAAAGGA 2297
QY 834 pProGluProLysTyrThrPheIleValIleGlnLysAtqHisAsnThrArgLeuLeuAr 854
Db 2298 CTACCGACCTGGGATCATTTATATTGTGGTGAGAAACGCATCACACCGCTTTTC-- 2355
QY 854 gArgMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspVa 874
Db 2356 -TGCTGTGACAAAGATGAGCGAATTGGGAAGAGTGGTAACATCCAGCT-----2403
QY 874 lAlaValAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyI 894
Db 2403 -----2403
QY 894 eValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAs 914
Db 2404 -----GGGACCACTGGGACACCAACATCCACCCATTTGAGTTGA 2447
QY 914 pPhePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrVa 934
Db 2448 CTTCATCTGTGAGCCAGCAGCATCCAGGCGCAGCCGACCATCCCATTTACTATGT 2507
QY 934 lMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAl 954
Db 2508 TCTTTGGGATGACAACCGTTTTCACAGCAGATGAGTCCAGATCCTGACGTACCAGCTGTG 2567
QY 954 aPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHi 974
Db 2568 CCACACTTACGTACGATGACAGCCTGTCTCTATCCAGACCTGCTCTACTATGCCCG 2627
QY 974 sLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAs 994
Db 2628 CCTGGTGGCTTTCCGGGACGA---TACCACCTGGTGGACAGAGCATCAGCTGGAGA 2684
QY 994 P 994
Db 2685 G 2685
```

## RESULT 2

```
US-09-949-016-4453
; Sequence 4453, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4453
; LENGTH: 7478
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4453
```

## Alignment Scores:

```
Pred. No.: 1,17e-55 Length: 7478
Score: 577.00 Matches: 237
Percent Similarity: 40.2% Conservative: 173
Best Local Similarity: 23.2% Mismatches: 370
Query Match: 10.8% Indels: 243
DB: Gaps: 38
```

US-10-645-746-3 (1-1020) x US-09-949-016-4453 (1-7478)

```
QY 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGlu-----37
Db 209 ATGGGATGGA-CGGGACCTCGGAGACGCTTACCTGCGGCGCTTACCTGCGGCGCTGAG 267
QY 38 -----LysLysValLeuLeu 42
Db 268 CAGGTGTTCCAGCACCTCGCGCGCTGGCATTGGCATTGGGGAACAACCAATCAAGCTC 327
QY 43 LeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrGluTyrGlu 62
Db 328 CTGGCCAATTTACTTTGAGGTGGGACATCCCTTAAGATCGACGCTGTACCCTACGAGGTG 387
QY 63 ValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysThr 82
Db 388 ATCAAGCCGGATGAG-----TGTCCTCCGTAGAGTC 417
QY 83 GluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGluLys 102
Db 418 AACCGGGAAGTGGTGAA-----TACATGCTCCAGCATTTCAAGCCT 459
QY 103 GlnThrAspPheIleLeuGluAspTyr-----ValPheAspGlyLysAspThrValTyr 120
Db 460 CAG-----ATCTTTGGTATCGCAAGCCTGTGTATGATCGAAGAAAGAAACAATTAC 510
QY 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysVal-----138
Db 511 -----ACTGTCACGACACTCCCATTTGGCAACGACGGGTGACTTT 552
QY 139 ---ValLysLysAspSerGlyLysLysAspGluLysAspLeuGlyLysIleLeuTyr 157
Db 553 GAGGTGACAATCCCTGGGGAAGGAGGAT---CGAATCTTTAAGGTCTCCATCAAGTGG 609
QY 158 ThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGlu 177
Db 610 CTAGCCATTGTGAGCTGGCGAATGTCATGAGGCCCTGTGTGTCAGCGCCAGATCCCTGTT 669
QY 178 LysAspGluGluAlaAsnArgSerTyrLysPheLysAsnValMetThrGlnLysVal 197
Db 670 CCTTGGAGTCTGTGCAAGCCCTGGATGTGGCCATGAGGACCTGGCATCC-----ATG 723
QY 198 ArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyr 217
Db 724 AGGTACACCCCTGTG-----738
QY 218 AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln 237
```

Db	739	---	GGCGCGCTCTCTTCTCACCGCGCTGAGGGCTACTACCAACCG-	-----	780
Qy	238	SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyIleLysGlu	257		
Db	781	-----	CTGGGGGTGGCGCGAGTCTGGTTCCGCTTACCAAGTCTGTGGCGCT	831	
Qy	258	LeuPheAspGlyGluProValLeuAsnPheAlaValAspLysLeuPheTyrAsnAla	277		
Db	832	GCCATG--	TGGAAGATGATGCTCAACATTCATGCTCAGCCACTGCCTTTTATAAGGCA	888	
Qy	278	ProLysMetSerLeuLeuAspTyrLeuLeuLeuLeuValAspProGlnSerCysAsnAsp	297		
Db	889	-----	CAGCCAGGTGATTGAGTCTTATGTCAGGTGTGGGACATCAGGAACATAGATGAG	942	
Qy	298	AspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAla	317		
Db	943	CAGCCCAAG-----	CCCTCACGGACTCT	966	
Qy	318	AlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLys-----	CysAla	334	
Db	967	CAGCGCGTTCGCTTACCACAGGAGATCAAGCGCCTGAAGGTGGAGTCA	1026		
Qy	335	GluValTrpAspAsn-----	GluMetSerArgLeuThrGluArgHisLeuThr	350	
Db	1027	CAGATGAAGAGAGAAAGTACCGCGTGTGTAATGTTACCGCTGCCCTGATGCCATCAGACA	1086		
Qy	351	PheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLysSerAspArg	370		
Db	1087	TTCCCTTACAGCTGGAGGTGGACAGCTGTG-----	1119		
Qy	371	GlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGluGluAsnLysLys	390		
Db	1120	-----	GAGTGCACAGTGGCACAGTATTTTCAAGCAGAAATATAAC	1158	
Qy	391	Phe---IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTyr	409		
Db	1159	CTTCAGCTCAAGTATCCCATCTGCCCTGCCCTACAGTTGGCCAGGAACAAAAGCATACC	1218		
Qy	410	AlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArgIle	429		
Db	1219	TACCTTCCCTAGAGGCTGTGAACATT---	GTGCGTGGCGAGCGCTGTATTAAAGCTG	1275	
Qy	430	AspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLys	449		
Db	1276	ACCAGCAACCAAGACTCGACCATGATAAGGCCACAGCTAGATCGCTCCAGACAGACAG	1335		
Qy	450	GluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheVal	469		
Db	1336	GAGGAGATCAGTCGCCTGATGAAGAATGCCACACTACAACTTAGATCC-----	TACATC	1389	
Qy	470	GluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys	489		
Db	1390	CAGGAATTTGGGATCAAAGTGAAGGATGACATGACGAGGTGACAGGGCGAGTCTGCCG	1449		
Qy	490	GluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro-----	505		
Db	1450	CGCGCCATCTGCAGTACGGCGCGGACCGGGCCATTGCCACACCAATCAGGTGTC	1509		
Qy	506	---ValIleArgGlyPheGlnGluLeuLeu---AsnValValProGluLysGluLeu	523		
Db	1510	TGGGACATCGGGG-----	AAACAGTTCTACAATGGGATGGATCAAGTCTGG	1560	
Qy	524	CysCysAlaValPheValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp	543		
Db	1561	GCCATCGCGCTTCGCAACCCCAAAA-----	CAGTGTGAGAAGAG-----	1602	
Qy	544	ValValLysPheTyrThrGluLeuIle-Gly-----	GlyCysLysPhe	557	
Db	1603	GTGCTCAGAACCTTCACAGACCACTGCGGAAGATTTTCCAAGGATCGGGGATGCCTATC	1662		
Qy	557	eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAl	577		
Db	1663	CAGGTCACCTGTTGTTCTGAATATGTCACAGGGGGCA-GACAGCTG-----	1710		

Qy	577	a	Thr	Lys	Asn	Glu	Tyr	Ala	Phe	Tyr	Lys	Asn	Cys	Thr	Leu	Asn	Thr	Gly	Ile	Gly	Arg	Phe	597	
Db	1710	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1710	
Qy	597	e	Glu	Ile	Ala	Ala	Thr	Glu	Ala	Ala	Val	Asn	Met	Phe	Glu	Arg	Leu	Pro	Asp	Lys	Glu	Gln	Ly	617
Db	1711	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1711	
Qy	617	s	Val	Leu	Met	Phe	Ile	Ile	Ile	Ser	Lys	Arg	Gln	Leu	Asn	Ala	Tyr	Gly	Phe	Val	Lys	His	637	
Db	1749	G	CT	G	C	A	G	C	T	C	A	T	T	G	T	C	A	T	C	T	G	C	A	1808
Qy	637	s	Tyr	Cys	Asp	His	Thr	Ile	Gly	Val	Ala	Asn	Gln	His	Ile	Thr	Ser	Glu	Thr	Val	Thr	Ly	657	
Db	1809	T	G	T	G	G	A	G	A	T	A	C	T	C	T	T	G	G	A	T	G	C	A	1868
Qy	657	s	Ala	Leu	Ala	Ser	Leu	Arg	His	Glu	Lys	Gly	Ser	Lys	Arg	Ile	Phe	Tyr	Gln	Ile	Ala	Le	677	
Db	1869	G	A	C	C	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1901	
Qy	677	u	Lys	Ile	Asn	Ala	Lys	Leu	Gly	Ile	Asn	Gln	Glu	Leu	Asp	Trp	Ser	Glu	Ile	Ala	Gln	Al	697	
Db	1902	C	A	A	G	A	T	C	A	A	T	T	G	T	G	C	A	T	T	A	C	A	1941	
Qy	697	u	Le	Ser	Pro	Glu	Glu	Lys	Glu	Arg	Gly	S	Thr	Met	Pro	Leu	Thr	Met	Tyr	Val	Gly	Ile	717	
Db	1942	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1994		
Qy	717	e	Asp	Val	Thr	His	Pro	Thr	Ser	Tyr	Ser	Gly	Ile	Asp	Tyr	Ser	Ile	Ala	Ala	Val	Val	Al	737	
Db	1995	A	G	A	T	G	T	T	A	C	A	C	C	C	C	C	C	C	C	C	C	C	2054	
Qy	737	a	Ser	Ile	Asn	Pro	Gly	Gly	Thr	Ile	Tyr	Arg	Asn	Met	Ile	Val	Thr	Gln	Glu	Cys	Arg	757		
Db	2055	C	A	G	A	T	G	A	T	G	C	C	C	C	C	C	C	C	C	C	C	C	2103	
Qy	757	g	Pro	Gly	Glu	Arg	Ala	Val	Ala	His	Gly	Arg	Glu	Arg	Thr	Asp	Ile	Leu	Glu	-----	-----	-----	2175	
Db	2104	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2135		
Qy	775	a	Lys	Phe	Val	Lys	-----	Leu	Leu	Arg	Glu	Phe	Ala	Glu	Asn	Asn	Asp	Asn	Arg	Ala	Pro	Al	794	
Db	2136	C	T	A	C	A	T	G	T	G	C	G	A	G	T	C	C	T	C	A	T	C	A	2192
Qy	794	a	His	Ile	Val	Val	Tyr	Arg	Asp	Gly	Val	Ser	Asp	Ser	Glu	Met	Leu	Arg	Val	Ser	His	As	814	
Db	2193	C	C	G	A	T	C	A	T	T	C	T	A	C	G	A	T	G	C	C	C	C	A	2252
Qy	814	p	Glu	Leu	Arg	Ser	Leu	Lys	Ser	Glu	Val	Lys	Gln	Phe	Met	Ser	Glu	Arg	Asp	Gly	Glu	As	834	
Db	2253	T	G	A	G	T	A	C	T	G	C	C	A	T	C	A	A	T	G	-----	-----	-----	2297	
Qy	834	p	Pro	Glu	Pro	Lys	Tyr	Thr	Phe	Ile	Val	Ile	Gln	Lys	Arg	G	His	Asn	Thr	Arg	Leu	Leu	854	

QY 934 lMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAl 954  
 Db 2508 TCITTTGGATGACCAACGTTTCACAGACAGATGCTCAGATCCTGACGTACACGCTGTG 2567  
 QY 954 aPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHi 974  
 Db 2568 CCACACTTACGTACGATGCACACGCTCTGTCTCTATCCAGACACCTGCCTACTATGCCG 2627  
 QY 974 sLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAs 994  
 Db 2628 CCTGGTGGCTTTCGGGCACGA---TACCACCTGGTGACCAAGGAGCATGACAGTGGAGA 2684  
 QY 994 p 994  
 Db 2685 G 2685

## RESULT 3

US-09-533-029-105

; Sequence 105, Application US/09533029

; Patent No. 6664446

; GENERAL INFORMATION:

; APPLICANT: Heard, Jacqueline

; APPLICANT: Broun, Pierre

; APPLICANT: Riechmann, Jose-Luis

; APPLICANT: Keddie, James

; APPLICANT: Pineda, Omaira

; APPLICANT: Adam, Luc

; APPLICANT: Samaha, Raymond

; APPLICANT: Zhang, James

; APPLICANT: Yu, Guo-Liang

; APPLICANT: Ratcliffe, Oliver

; APPLICANT: Pilgrim, Marsha

; APPLICANT: Jiang, Cai-Zhong

; APPLICANT: Reuber, Lynne

; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES

; FILE REFERENCE: MBI-010

; CURRENT APPLICATION NUMBER: US/09/533,029

; CURRENT FILING DATE: 2000-03-22

; EARLIER APPLICATION NUMBER: 60/125,814

; EARLIER FILING DATE: 1999-03-23

; NUMBER OF SEQ ID NOS: 121

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 105

; LENGTH: 2910

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; OTHER INFORMATION: G1149

; US-09-533-029-105

## Alignment Scores:

Pred. No.:	3.39e-56	Length:	2910
Score:	575.50	Matches:	237
Percent Similarity:	39.0%	Conservative:	159
Best Local Similarity:	23.3%	Mismatches:	337
Query Match:	10.8%	Indels:	283
DB:	3	Gaps:	41

US-10-645-746-3 (1-1020) x US-09-533-029-105 (1-2910)

QY 38 LysLysValLeuLeuLeuValAlaGlnTyrPheLysPheSerSerLysIleTyrAspArgGlu 57  
 Db 472 AAGAAAGTCATGGTTCGTGGCAAT-----CATTTCTTGGTTCAGGTTCGTGATCGTAT 525  
 QY 58 TyrTyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysPro 77  
 Db 526 CTCTACCATTCAGATGTTTCGATCAATCCTGAGGTTATATCAAG----- 570  
 QY 78 PheProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeu 97  
 Db 570 ----- 570  
 QY 98 ArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAsp 117

Db 570 ----- 570  
 QY 118 ThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLys 137  
 Db 571 -----ACAGTGAACAGAAACGTGATGAACCTTCCTGGTTAAGAT--- 609  
 QY 138 ValValLysLysAspSerGlu-----LysLysAspGluLysAspLeuGluLysLys 154  
 Db 610 -----TATAAAGACTCTCACTTGGGAGGGAAGTCACAGCGTATGAT---GGAAGGAA 660  
 QY 155 IleLeuTyrThrMet-----IleLeuTyrArgLysLysPheHisLeuAsnPheSer 172  
 Db 661 AGCCTCTATACTGCTGCTGCTCATTACCTTTTGACCTCGAAAGAGTTGTGTGAATCTGGCG 720  
 QY 173 ArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyrLysPhe---LeuLysAsn 191  
 Db 721 GAGAAAGAGCTGACGGTTCCTCTGGGAAGGACAGACCGTTTAAAGTTGCTGTGAAGAT 780  
 QY 192 Val-----MetThrGlnLysValArgTyrAla 200  
 Db 781 GTGACCAAGCACTGATCTTTATCAGTTGCCAACAGTTCCTTGATCGTAAGCAAGAGAGGCT 840  
 QY 201 ProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyrAspAsn 220  
 Db 841 CCATAT-----GATACTATCAAGTCTTGATGTTGTTCTTAGGGATAAGCCCTCTAAT 894  
 QY 221 SerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGlnSerLeu--- 239  
 Db 895 GATTATGCTCTGTTGGAGGTCCTTT-----TTCACACTAGTTTGGGA 939  
 QY 240 -----GluValAlaProArgIleGluAlaTyrPheGlyIle 251  
 Db 940 AAGGACGCAAGAGATGTTAGGGGTGAGCTTGGAGATGGTATTGAGTACTGGAGAGTTAT 999  
 QY 252 TyrIleGlyIleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAsp 271  
 Db 1000 TTCCAAAGTCTAAGG---CTGACTCAGATGGGTTTGTCTCTGAACACTGACGTTTCAGCA 1056  
 QY 272 LysLeuPheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuValAsp 291  
 Db 1057 AGATCATTTTAT-----GAA 1071  
 QY 292 ProGlnSerCysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLys 311  
 Db 1072 CCGATTGTTGTCACCTGACTTTATTAGCAAGTTTCTGAATATAAGGGACTTA-----AAC 1125  
 QY 312 MetThrIleArgGlnAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeu 331  
 Db 1126 AGGCCACTTAGAGACTCAGATCGACTTAAGGTGAAGAAAGTTTGTGAGGACACTGAAAGTT 1185  
 QY 332 LysCysAlaGluValTyrAspAsnGluMetSerArg----- 343  
 Db 1186 AGTTGCTTCAC---TGGAAACGACACAAAAGTGCACAAATATTAGTGGGATTTCTAGTCTA 1242  
 QY 344 ---LeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyr 362  
 Db 1243 CCCATCAGGAGCTAAGGTTCACTCTGGAGGACAAATCAGAGAAGACGGTTGTTCAATAT 1302  
 QY 363 LysValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPhe 382  
 Db 1303 -----TTTGTGAAAAATATAAT----- 1320  
 QY 383 LysIleTyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysVal 402  
 Db 1321 -----TATAGA-----GTGAATACAGGCTCTACTGCTGATTTCAACA 1359  
 QY 403 LysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysPro 422  
 Db 1360 GGGAGGTGACAAACACCGCTTACCTTACCAATGGAGCTCTGCCAAATTCAGGAA---GGG 1416  
 QY 423 GlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr 442

```

Db      1417 CAAGATACACCAAGGCTCAATGAGAACCAAGTGCATGCTGCTAAAGACTACCTGC 1476
QY      443 ArgLysProHisAspTyrLysGluAenThrLeuLysMetLeuLysGluLeuAaspPheSer 462
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1477 CAACGACCCCTGAT---AGAGAGAACTCGATCAAAAACCTGGTTGTGAAAAATAATTAC 1533
QY      463 SerGluGluLeuAenPheValGluArgPheGlyLeuLysSerLysLeuGlnMetIleGlu 482
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1534 AATGATGATCTGAGC-----AAGGAGTTTGGGATGTCAGTCACTACCCCACTAGCCTCG 1587
QY      483 CysProGlyLysValLeuLysGluProMetLeuValAenSerValAenGluGlnIleLys 502
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1588 ATTGAAGCTCGTGACTTCCGCCACCGATGTTGAAGTACCATGACAGTGTAAAGAGAAA 1647
QY      503 MetThrProValIleArgGlyPheGlnGluLysGlnLeuAenValValProGluLysGlu 522
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1648 ATGGTAAATCCAAGGCTAGGA-----CAGTGAACATGATTGACCAAGAAA--- 1692
QY      523 LeuCysCysAlaValPheValAenGluThrAlaGlyAenProCysLeuGluGluAen 542
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1692 ----- 1692
QY      543 AspValValLysPheTyrThrGluLeuIleGlyCysLysPheArgGlyIleArgIle 562
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1693 -----ATGGTTAATGGAGCAAA-----GTCACCTTCT 1719
QY      563 GlyAlaAenGluAenArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAenGluTyr 582
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1720 TGGACTTGGGAATTTAAGCCTCAACCTGCTATTTCGGTTTCATCTTGTCCCTCGAACAT 1779
QY      583 AlaPheTyrLysAenCysThrLeuAenThrGlyIleGlyArgPheGluIleAlaAlaThr 602
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1780 -----ATTGGAAGCTCTTCTC 1797
QY      603 GluAlaLysAenMetPheGluArgLeuProAaspLysGluGlnLysValLeuMetPheIle 622
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1798 GAT-----ATCCACAAAGGGCACCTGTCTCCAA-----CTGTGATT 1836
QY      623 IleIleSerLysArgGlnLeuAenAlaTyrGlyPheValLysHisTyrCysAaspHisThr 642
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1837 GTAATATTCCCTGATGTCATCATATGGAATAAATAAAGGATCTGTGAAACAGAA 1896
QY      643 IleGlyValAlaAenGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeu 662
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1897 TTGGGATTTGCTCTCAGTGTGCCAACCTAGACAAAGTTAATAAATC----- 1944
QY      663 ArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAenAlaLys 682
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1945 -----AACAAGCAGTACATGGAATAATGTTGCCTTGAAGATCAATGTCAAG 1989
QY      683 LeuGlyGlyIleAenGlnGluLeuAaspTrpSerGluIleAlaGluLysSerProGluGlu 702
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1990 ACTGGGGGAAGGAACACTGTTCTTAT-----ThrMetTyrValGlyIle 717
QY      703 LysGluArgArgLysThrMetProLeu-----ThrMetTyrValGlyIle 717
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2017 GATGCTATTAGAAGAAACAACTCTTATTACTGATCGTCCAAACCAATCATCATGGTGTCT 2076
QY      718 AspValThrHisProThrSerTyrSerGlyIleAaspTyrSerIleAlaAlaValValAla 737
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2077 GATGTGACTCACCCACAGCTGGAGAGACTCAAGTCCCTTCTATTGCTGCTGTGTGGCC 2136
QY      738 SerIleAen---ProGlyGlyThrIleTyrArgAenMetIleValThrGln----- 753
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2137 TCTATGAGTGGCCTGAGATAAACAAATACCGAGGATTTGGTTTCTGCTCAAGCTCATAGG 2196
QY      754 GluGlu-----CysArgProGlyGluArgAlaValAlaHis 765
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2197 GAAGAAATTTATCAGACCTGATTAAGCTGGTTTCAGGATCCCAACGTCGGGTAGTCCAC 2256
QY      766 GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla 785
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2257 -----TCTGGTTTGATAGGGAACATTTTCATGCCATTCAGGAGAGCT----- 2298
```

```

QY      786 GluAenAenAaspAenArgAlaProAlaHisIleValValTyrArgAaspGlyValSerAasp 805
Db      2299 -----ACAGGCCAGATACCTCAAGGATCATCTTCTATCGTGACGGAGTAAGCGAA 2349
QY      806 SerGluMetLeuArgValSerHisAaspGluLeuAargSerLeuLysSerGluValLysGln 825
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2350 GGGCAGTTTAGTCAGGTTCTGCTACATGAGATGACTGTATCCGCAAGCCTTGTAACCTCT 2409
QY      826 PheMetSerGluArgAaspGlyGluAaspProGluProLysTyrThrPheIleValIleGln 845
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2410 CTCCAA-----GAGAAATTATGTTCTCGTGTTACTTTCGTGATTGTCCAG 2454
QY      846 LysArgHisAenThrArgLeuLeuAargMetGluLysAaspLysAaspValValAenLys 865
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2455 AAACGTCACACACACGTTTGTTCCTCAGCAACACGGG-----AATCGT 2499
QY      866 AspLeuThrProAlaGluThrAaspValAlaValAlaAlaValLysGlnTrpGluGluAasp 885
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2500 GATATGACT----- 2508
QY      886 MetLysGluSerLysGluThrGlyIleValAenProSerSerGlyThrThrValAaspLys 905
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2509 -----GATAAGAGTGGCAATATTCAACCA-----GGTACTGTCGTGGACACT 2550
QY      906 LeuIleValSerLysTyrLysPheAaspPhePheLeuAlaSerHisGlyValLeuGly 925
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2551 AAAATCTCTCACCTTAATGAATTCGATCTTATTTGAACAGCCTGCTGTTATTCAGGA 2610
QY      926 ThrSerArgProGlyHisTyrThrValMetTyrAaspAaspLysGlyMetSerGlnAaspGlu 945
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2611 ACNAGCAGCGGCACATTTACCATGTTCTCTCGATGAGACGGTTTCACCGCTGATCAG 2670
QY      946 ValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSer 965
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2671 TTGCAATGCTCACAAACAACTCTGCTACAGCTATGCGAGGTGTACAAAATCTGTGTCA 2730
QY      966 LeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArg 985
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2731 ATTGGCCACCAACGCTACTACGCTCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2778
QY      986 ThrTyrLysGluHisTyrIleGlyAaspTyrAlaGlnProArgThrArg 1001
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2779 TACTACATGAGAGTGGATGCTCTGATGGAGGTTTCGAGCAGGTCCAGG 2826

RESULT 4
US-09-270-767-11928
; Sequence 11928, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11928
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11928

Alignment Scores:
Pred. No.: 3,348-42 Length: 1662
Score: 452.50 Matches: 156
Percent Similarity: 43.8% Conservative: 91
Best Local Similarity: 27.7% Mismatches: 213
Query Match: 8.5% Indels: 104
DB: 3 Gaps: 18

US-10-645-746-3 (1-1020) x US-09-270-767-11928 (1-1662)
```



Db 182 GTATACCGTGAATGGTCTTTCTCGTGGCCCGCCAGCAGTGAGACCTTCGAAACACGAT 241  
 Qy 379 -----ThrThrLeuPheLysIleTyrGluGluAsnLysPheIleGluPhePro 395  
 Db 242 GGGAGAAAGGTACCATTCGCTCTACTTCCACAGTCGCAACTACCCATTGAAGTTTCC 301  
 Qy 396 HisLeuProLeuValLysSerGlyAlaLysGluTyrAlaValProMetGluHis 415  
 Db 302 CAATCCCATGTGTGAACGTTCGATCTTCAATCAAGAGTATTCTCGTCCCATCAGGTA 361  
 Qy 416 LeuGluValHisGluLysProGlnArgTyrLysAsnArgIleAspLeuValMetGln--- 434  
 Db 362 TCGAGCATCGAGAA-----GGTCAGGCCCTAAACCGCAAGGATCGAGCAACTCAGGTG 415  
 Qy 435 AspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLysGluAsnThrLeuLys 454  
 Db 416 GCCAATATGATAAGATACGACGACCATCGAGCAACGTCGGAAGCGCAAGATATGAAC 475  
 Qy 455 MetLeuLysGluLeuAspPhe-SerSerGluGluLeuAsnPheValGluArgPheGlyLe 474  
 Db 476 TTGCTGCAATAC-----TTCCAGCACAAACCTGGGATCCGACCATCAGTCGCTTTGGCAT 529  
 Qy 474 uCySerLysLeuGlnMetIleGluCysProGlyLysValLeuLysGluProMetLeuVa 494  
 Db 530 CGCATTCGCAACGATTTATTGTGGTAAGCACCCCGCTCTAAACCCACCTCAGGT-- 587  
 Qy 494 IAsnSerValAsnGluGlnIleLysMetThrProValIleArgLysPheGlnGlu----- 512  
 Db 588 -----GAATATCATAGTAAGAGGTTTACTATGTGTGAAGACGGGTCTGCGCGCATGGA 640  
 Qy 513 -----LysGlnLeuAsnValProGluLysGluLeuCysCysAlaValPheVal-- 529  
 Db 641 TGGCATGAAGTTTCTGGAGCCCAAGCCCAAGGCGCACAGTGTGCGGTCTGTATTTGCGA 700  
 Qy 530 -----ValAsnGluThrAlaGlyAsnProCy 538  
 Db 701 TCCGAGGAGTGTGCGAAATGAACCTATGATCCCGAGCTGAATGAC---TTGCGGAACCTAAT 757  
 Qy 538 sLeuGluGluAsnAspValValLysPheTyrThrGluLeuIleGlyGlyCysLysPheAr 558  
 Db 758 AATATCCCAAGGCAAGGCGTCAACATA-----ACCTTGGATTCTGATGTGCATACAG 811  
 Qy 558 gGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAlaTh 578  
 Db 812 ACCGTT-----ACGGATCGAACCGACCGCTAGACATATTTCGCGATCTGMA 862  
 Qy 578 rLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGl 598  
 Db 863 GCGCAGCCAGCAGAT----- 878  
 Qy 598 uIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLysVa 618  
 Db 878 ----- 878  
 Qy 618 lLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTy 638  
 Db 879 -CTGCAATTGTGATATTCTCAGTTAGATT---TCTACGATACAAATTAAACAGAA 934  
 Qy 638 rCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAl 658  
 Db 935 GCGCAGCTGCAGCATGGAATTTTGACGCAATGCATTAAAGCAGTTCACCGTGAACGA-- 992  
 Qy 658 leuAlaserLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLys 678  
 Db 993 -----AAGTGTAAATATCAGACGATTTGGAATAATTCTACTTAA 1030  
 Qy 678 sIleAsnAlaLysLeuGlyIleAsnGlnGluLeuAspTyrSerGluIleAlaGluI 698  
 Db 1031 GATCAACTCAAGCTGAACGGGATCAACCAAGATC----- 1067  
 Qy 698 eSerProGluGluLysGluArgLysThrMetProLeuThrMetTyrValGlyIleAs 718  
 Db 1068 -----AAGGATGATCTCTCGTATGATGAAGAACACCATGATATTGGAGCCGA 1120

Qy 718 pValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValValAlase 738  
 Db 1121 TGTGACCATCCCTCTCCGATCAGCGGAGATTCCAGTGTGGTGGGATGACAGCCTC 1180  
 Qy 738 rIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGluCysArgPr 758  
 Db 1181 ACAGCATCCTACGAGGCGCAGTTAATCAATGCAATATCGTTTGCAG----- 1226  
 Qy 758 oGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPhe-- 777  
 Db 1227 -----CGAGGGCTCTCGAGGAGATTGAGACATGTTCTCTC 1261  
 Qy 778 -----ValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAlaH 795  
 Db 1262 GATTACTTTGGAGCACTTGGCGTGTATAAGAGTACCGT---AACGCCATCTCTGATCA 1318  
 Qy 795 sIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspGl 815  
 Db 1319 TATCATCTACTACGAGATGGCGTGAGCGCGCCAGTTTCGAAAAATCAAAAACGAGGA 1378  
 Qy 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835  
 Db 1379 ACTGAGGTGTATTAACAAGCCTGT-----GACAAGGTGGGCTG 1417  
 Qy 835 oGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgAr 855  
 Db 1418 TAAACCCAGGATTTCTGCGTATTTGGTGAAGCGTCATCACCTCTCTTCTTCCC-- 1475  
 Qy 855 gMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAl 875  
 Db 1475 ----- 1475  
 Qy 875 aValAlaValLysGlnTrpGluAspMetLysGluSerLysGluThrGlyIleVa 895  
 Db 1476 -----AGCGGCGACGTAACGACATCGAAACAAGTTTCAACAACGT 1513  
 Qy 895 lAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPh 915  
 Db 1514 GGACCCC-----GGACCCGTGTCGATCGCACCATTTGTCATCTCAACGATGATGAT 1567  
 Qy 915 ePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935  
 Db 1568 CTTTCATGTGACGCCACCGGATCCAGGCGACGCGCAAGCCACACGATACAAATGTGAT 1627  
 Qy 935 tTyrAspLysGlyMetSerGluAspGluValTyrLysMetThrTyrGlyLeuAlaph 955  
 Db 1628 TCAGAACACAGGCAATCTTGACATCGACTTGTTCGACGAGTTGACCTACAACCTGTGCCA 1687  
 Qy 955 eLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHisLe 975  
 Db 1688 CATGTTCTCTCGTTGCAATCGCTCGTTTCTTATCGGCTCGGCTATTAGCCCATTT 1747  
 Qy 975 uSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIleGly----- 993  
 Db 1748 GGTACCTGCTCGTGACGC---GTTTATCTGATCGCACCAACAGGTTCTCTGATTTGAA 1804  
 Qy 994 -----AspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnThrAsnVa 1012  
 Db 1805 GAAGGAGTACGCAAG-----CGAACGATTTGTCGCCGAATTCATGAAGAAAAACCC 1855  
 Qy 1012 lLysTyr 1014  
 Db 1856 CATGTAC 1862

## RESULT 6

US-10-043-774B-1  
 ; Sequence 1, Application US/10043774B  
 ; Patent No. 690017  
 ; GENERAL INFORMATION:  
 ; APPLICANT: University of Illinois at Chicago  
 ; APPLICANT: Sharma, Arun  
 ; APPLICANT: Hoffman, Ronald



; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES

; FILE REFERENCE: MBHB: CU08/PPA  
; CURRENT APPLICATION NUMBER: US/10/043,774B  
; CURRENT FILING DATE: 2002-06-06  
; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 2328

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(2328)

; OTHER INFORMATION: Human Hiwi Protein

US-10-043-774B-1

Alignment Scores:

Pred. No.: 1,78e-25 Length: 2328

Score: 312.00 Matches: 150

Percent Similarity: 36.7% Conservative: 88

Best Local Similarity: 23.1% Mismatches: 218

Query Match: 5.8% Indels: 192

DB: 3 Gaps: 27

US-10-645-746-3 (1-1020) x US-10-043-774B-1 (1-2328)

Qy 417 GluValHisGlu---LysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAsp 435  
Db 613 GAAGAATAAATTTCAAGAACAAAGTTTCCAAAGAACTAATAGTTTGTCTTACC 672  
Qy 436 LysPheLeuLysArgAlaThrArgLysProHis---AspTyrLysGluAsnThrLeuLys 454  
Db 673 AAGTATAACATTAAGACATACAGATGGATGATATTGACTGGGACCAAGAAATCCCAAGAGC 732  
Qy 455 MetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheValGluArgPheGlyLeu 474  
Db 733 ACCTTTAAGAAAGCCGAGGCTCT-----GAAGTCAGCTTCTTAGAATACTACAGGAAG 786  
Qy 475 CysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLysGluProMetLeuVal 494  
Db 787 CAATACAACCAAGAGATACCCGAC-----TTGAAGCAGCCTGTCTTGGTC 831  
Qy 495 AsnSerValAsnGluGln-----Lys-----Ile 501  
Db 832 AGCCAGCCCAAGAGAGGGGGCCCTTGGGGGACACTGCGAGGGCTGCCATGCTCAT 891  
Qy 502 LysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsn----- 516  
Db 892 CCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGCGTAATGATTTTAACGTGATG 951  
Qy 517 -----ValValProGluLysGluLeuCysCysAlaValPhe 528  
Db 952 AAAGACTTACCGCTTTCATCAAGACTAACTCCAGAGCAAAAG----- 993  
Qy 529 ValValAsnGluThrAlaGlyAsn-----ProCysLeuGluGluAsnAspValValLys 546  
Db 994 -----CAGCGTGAAGTGGGACGACTCATTGATTACATCAATAAAACGATAATGTTCAA 1047  
Qy 546 ----- 546  
Db 1048 AGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCCACTTACTGTCTCTCAGGAAGA 1107  
Qy 547 ----PheTyrThrGluLeuLe-----GlyGlyCysLysPhe-----ArgGlyIle 560  
Db 1108 ATTTTGCACAAACAGAAAGATTTTCAACCAAGGTGGGAAAAACATTTGATTACAATCCACAATTT 1167  
Qy 561 ArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAsn 580  
Db 1168 GCAGATTGCTCCAAAGAAACAAGAGGTGCACCAATTAATTAGTTTAAAGCCACTAGATAAC 1227  
Qy 581 GluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAla 600  
Db 1228 TGGCTGTTGATCTATACGCGA-----AGAAATTATGAGCAGGCC 1266

Qy 601 AlaThrGluAlaLysAsnMetPheGluArgLeuPro----- 612  
Db 1267 AATTCAATTGATACAAAATCTATTTAAAGTTTACACGACCATGGCATGCAAAATGAGAAA 1326  
Qy 613 -----AspLys-----GluGlnLys 617  
Db 1327 GCATAATGATTGAAGTGGATGACAGACGTGAAGCCCTACTTAAGAGTCTTACAGCAAAAG 1386  
Qy 618 Val-----LeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyr 632  
Db 1387 GTCACAGCAGACACCCAGATAGTTGTCTGTCTGTGTCAAGTAATCGGAGGCAAAATAC 1446  
Qy 633 GlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSer 652  
Db 1447 GATGCTATTAAAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAAGTGTGTGTGGCC 1506  
Qy 653 GluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePhe 672  
Db 1507 CGAACCTTTAGGCAAA-----CAGCAAACTGTCATGGCCATTTGCT 1545  
Qy 673 TyrGlnIleAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTrp 692  
Db 1546 ACAAGATTGCCCTACAGATGAACCTGCAAGATGGGGA-----GAGCTC---TGG 1593  
Qy 693 SerGluIleAlaGluIleSerProGluGluLysGluArgArgLysThrMetPro----- 710  
Db 1594 -----AGGTGGACATCCCTCCCTGAAG 1614  
Qy 711 LeuThrMetTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyr 730  
Db 1615 CTGCTGATGATCCTTGGCATGATTGTACCATGACATGACATGCTGGG-----CGAGG 1668  
Qy 731 SerIleAlaAlaValValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIle 750  
Db 1669 TCAATCGAGGATTTGTTGCCACATCAATGAAGGATGACCCGCTGTTCTCAGCTGC 1728  
Qy 751 ValThrGlnGluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThr 770  
Db 1729 ATATTTTCAGAT-----AGAGGACAGGAGCTGCTA 1758  
Qy 771 AspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAspAsn 790  
Db 1759 GATGGGCTCAAGCTCGCTCCGCAAGCGGCTCTGAGGGCTTGGAAATAGCTGCAATGAG--- 1815  
Qy 791 ArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArg 810  
Db 1816 TACATGCCAGCCGCGATCATCGTGTACCGCGATGGCGTAGGAGACGCGCCAGCTGAAACA 1875  
Qy 811 ValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSer----- 828  
Db 1876 CTG-----GTGAACCTACGAAGTGCACAGTTTTTTGGATTGTCTA 1914  
Qy 829 GluArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHis 848  
Db 1915 AAATCAATTGGTAGAGTTTACACCTAGACTAACCGTAAATTGTGTGTGAAGAAAGAGTG 1974  
Qy 849 AsnThrArgLeuLeuArgMetGluLysAspLysProValValAsnLysAspLeuThr 868  
Db 1975 AACACCAAGATTTTT----- 1989  
Qy 869 ProAlaGluThrAspValAlaValAlaValLysGlnTrpGluGluAspMetLysGlu 888  
Db 1989 ----- 1989  
Qy 889 SerLysGluThrGly-----IleValAsnProSerSerGlyThrThrValAspLysLeu 906  
Db 1990 ---GCTCAGTCTGGAGGAAGACTTCAGAAATCCACTTCTCGGAAACAGATTATTGATGAGAG 2046  
Qy 907 IleValSerLysTyrLysPheAspPhePheLeuAlaSerHisHisGlyValLeuGlyThr 926  
Db 2047 GTTACCAGACCAAGATGTTGATGACTTTTTTTATTCGTGAGCGCAGCTGTGAGAGTGTAGT 2106

QY 927 SerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluVal 946  
Db 2107 GTTCTCCACACATTACAAATGTCTATGACACACGGCGCTGAAGCCAGCACACATA 2166  
QY 947 TyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeu 966  
Db 2167 CAGCGCTGACCTACAGCTGTGCCACATCTATTACAACTGGCCAGGTGCTATTGCTGTT 2226  
QY 967 ProValProValHisTyrAlaHis 974  
Db 2227 CCGTCTCTTGCAGTACGCCAC 2250

## RESULT 7

US-09-873-737A-5  
; Sequence 5, Application US/09873737A  
; Patent No. 6723534  
; GENERAL INFORMATION:  
; APPLICANT: Duke University  
; APPLICANT: Lin, Haifan  
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE  
; FILE REFERENCE: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME  
; CURRENT APPLICATION NUMBER: US/09/873,737A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: PCT/US99/28764  
; PRIOR FILING DATE: 1999-12-03  
; PRIOR APPLICATION NUMBER: 60/110,901  
; PRIOR FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 3472  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (167)..(2749)  
; NAME/KEY: misc.feature  
; LOCATION: (392)  
; OTHER INFORMATION: n=t or a, Xaa=Leu or Ile  
; NAME/KEY: misc.feature  
; LOCATION: (1073)  
; OTHER INFORMATION: n=c or a, Xaa=Leu or Ile  
; NAME/KEY: misc.feature  
; LOCATION: (2369)  
; OTHER INFORMATION: n=c or a, Xaa=Leu or Ile  
US-09-873-737A-5

Alignment Scores:  
Pred. No.: 1,81e-24 Length: 3472  
Score: 306.00 Matches: 151  
Percent Similarity: 37.4% Conservative: 88  
Best Local Similarity: 23.6% Mismatches: 226  
Query Match: 5.7% Indels: 174  
DB: 3 Gaps: 26

US-10-645-746-3 (1-1020) x US-09-873-737A-5 (1-3472)

QY 417 GluValHisGlu---LysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAsp 435  
Db 1037 GAAGACATAAATTTCAAGAACACAGTTTCCAAAGAANTATAGTTTACTGTTCTTACC 1096  
QY 436 LysPhe---LeuLysArgAlaThrArgLysProHisAspTyrLysGluAsnThrLeuLys 454  
Db 1097 AAGTATAACATTAAAGACATACAGAGTGGATGATATTGACTGGGACCAAGAATCCCAAGAGC 1156  
QY 455 MetLeuLysGluLeuAspPheSer----- 462  
Db 1157 ACCTTTAAGAAAGCCGACCGCTCTGGGGTCAGCTCTTAGAATACTACAGGAAGCAATAC 1216  
QY 463 SerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGlu 482  
Db 1217 AACCAAGAGATCACCAGCTTGAAGCAGCCCTGTCTTGGTCAGCCAGGCCCAAGAGAGCGG 1276

QY 483 CysProGlyLysValLeuLysGluPro---MetLeuValAsnSerValAsnGluGlnIle 501  
Db 1277 GSCCCTGGGGGACACCTGCAGGGCTGCCATCTCATTTCTGAGCTGCTATCTTACA 1336  
QY 502 LysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGluLys 521  
Db 1337 GGTCTAACTGATAAATGCGT-----AATGATTTTAACTGTATG-----AAA 1378  
QY 522 GluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCys----- 538  
Db 1379 GACTTAGCGGTTTCATACAGACTAATCCAGAGCAAGCGCAGCTGTAAGTGGGACGATC 1438  
QY 539 -----LeuGluGluAsnAspValValLys----- 546  
Db 1439 ATTGATTACATTTATAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGC 1498  
QY 547 -----PheTyrThrGluLeuIle----- 552  
Db 1499 TTTGATTCCAACTTACTGCTCTTCTCAGGAAGAATTTTCCAAACAGAAAAGATTACCAA 1558  
QY 553 GlyGlyCysLysPhe-----ArgGlyIleArgIleGlyAlaAsnGluAsnArgGly 569  
Db 1559 GGTGGAAGAAACATTTGATTACAATCCACAATTTGCAGATTGGTCCAAAGCAACAGAGGT 1618  
QY 570 AlaGlnSerIleMetTyrAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThr 589  
Db 1619 GCACCATTAATTAGTGTAAAGCCACTAGATAACTGGCTGTGTGATCTATATACGCGA----- 1672  
QY 590 LeuAsnThrGlyIleGlyArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGlu 609  
Db 1673 -----AGAAATATGAGCAGCCCAATTCATTGATACAAAATCTATTAA 1717  
QY 610 ArgLeuPro-----AspLys 614  
Db 1718 GTTACACCCAGCCATGGGCATGCAAAATGAGAAAAGCAATAATGATTGAAGTGGATGACAGA 1777  
QY 615 -----GluGlnLysVal-----LeuMetPhe 621  
Db 1778 ACTGAAGCCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTC 1837  
QY 622 IleIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAspHis 641  
Db 1838 TGTCTGTTGTCAGTAATCGGAGGACAAATACGATGCTATTAAAAAATACCTGTGTACA 1897  
QY 642 ThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSer 661  
Db 1898 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCAACCTTAGGCAAA----- 1945  
QY 662 LeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsnAla 681  
Db 1946 -----CAGCAAACTGTCTAGGCCATTTGTACAAAGATTGCCCTACAGATGAAGTGC 1996  
QY 682 LysLeuGlyGlyIleAsnGlnGluLeuAspTyrSerGluIleAlaGluIleSerProGlu 701  
Db 1997 AAGATGGGAGGA-----GAGCTC---TGG----- 2017  
QY 702 GluLysGluArgArgLysThrMetPro-----LeuThrMetTyrValGlyIleAspVal 719  
Db 2018 -----AGGGTGGACATCCCTCCCTGAAAGCTCGTGATGATCGTTGGCATCGATTGT 2065  
QY 720 ThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValValAlaSerIle 739  
Db 2066 TACCATGACATGACCGCTGGTTCTCAGCTGCATATTTTCAGAT----- 2119  
QY 740 AsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGluCysArgProGly 759  
Db 2120 AATGAAGGATGACCGCTGGTTCTCAGCTGCATATTTTCAGAT----- 2164  
QY 760 GluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLys 779  
Db 2165 -----AGAGGACAGAGCTGGTGTAGATGGGCTCAAAAGTCTGCTCGCAAGCG 2209







QY 893 yileValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPh 913  
Db 2737 ----GTAACCTCCACCTCCTGGAACTCTGTGTAGATCATACAAACAGCTGTGAGTGGT 2792  
QY 913 eAspPheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrTh 933  
Db 2793 GGATTTCTATCTTCTTGGCCCATCATGTACGCCAGGCTGTGGCATTCCTACGCATTATGT 2852  
QY 933 rValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLe 953  
Db 2853 CTGTGTTCTCAACACCGCAACCTGTAGCCCTGATCATATGCAGAGCTGACTTCAAACT 2912  
QY 953 uAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAl 973  
Db 2913 GTGCCACATGTACTGGAATTGGCTGGCCACCATCAGATTCCAGTCTCTTGAAGTATGC 2972  
QY 973 aHis 974  
Db 2973 CCAC 2976

RESULT 10  
US-10-104-047-1469  
; Sequence 1469, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cDNA  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1469  
; LENGTH: 2579  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1469

Alignment Scores:  
Pred. No.: 1,39e-22 Length: 2579  
Score: 288.00 Matches: 144  
Percent Similarity: 36.7% Conservative: 87  
Beat Local Similarity: 22.9% Mismatches: 207  
Query Match: 5.4% Indels: 132  
DB: 3 Gaps: 27

US-10-645-746-3 (1-1020) x US-10-104-047-1469 (1-2579)

QY 417 GluValHisGlu---LysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAsp 435  
Db 924 GAAGAACATAAATTTCAAGAACAAAGTTTCCAAAGAACTAATAGTTTGTGTCTTACC 983  
QY 436 LysPheLeuLysArgAlaThrArgLysProHis---AspTyrLysGluAsnThrLeuLys 454  
Db 984 AGATATAACAATAAGACATACAGAGTGGATGATATGACTGGACCAAGATCCCAAGACC 1043  
QY 455 MetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheValGluArgPheGlyLeu 474  
Db 1044 ACCTTTAAGAAAGCCGACGGCTCT-----GAAGTCAGCTTCTTAGAATACTACAGGAAG 1097  
QY 475 CysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLysGluProMetLeuVal 494  
Db 1098 CAATACAAACCAAGAGATACCCGAC-----TTGAAGCAGCCTGTCTTGGTC 1142  
QY 495 AsnSerValAsnGluGln-----ile 501  
Db 1143 ACCCAGCCCAAGAGAAGCGGGCCCTGGGGGACACTGCCAGGGCCTCCATGTCTATT 1202  
QY 502 LysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsn----- 516  
Db 1203 CCTGAGCTCTGCTATCTTACAGGTCTTAACGTATGATAAAATCGTAATGATTTTAAACGTGATG 1262

QY 517 -----ValValProGluLysGluLeuCysCysAlaValPhe 528  
Db 1263 AAAGACTTAGCGCTTCATACAAGACTAACTCCAGAGCAAAAG----- 1304  
QY 529 ValValAsnGluThrAlaGlyAsn-----ProCysLeuGluGluAsnAspValValLys 546  
Db 1305 -----CAGCGTGAAGTGGAGCACTCATTTGATTACATTCAATAAAACCAATATGTTCAA 1358  
QY 546 ----- 546  
Db 1359 AGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTTCCACTTGTCTTCTCAGGAAGA 1418  
QY 547 ---PheTyrThrGluLeuIle-----GlyGlyCysLysPhe-----ArgGlyIle 560  
Db 1419 ATTTTGGCAAAACAGAAAGATTCCCAAGGTGGAATAACATTGATTACATCCACCAATTT 1478  
QY 561 ArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAsn 580  
Db 1479 GCAGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTAAAGCCACTAGATAAC 1538  
QY 581 GluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAla 600  
Db 1539 TGGCTGTTGATCTATACCGCA-----AGAAATTATGAAGCAGCC 1577  
QY 601 AlaThrGluAlaLysAsnMetPheGluArgLeuPro----- 612  
Db 1578 AATTCATTGATACAAAAATCTATTAAAGTTACACCAGCCATGGCATGCAAAATGAGAAAA 1637  
QY 613 -----AspLys-----GluGlnLys 617  
Db 1638 GCAATAATGATTGAAGTGGATGACAGAACTGAAGCTACTTAAGAGTCTTTACAGCAAAAG 1697  
QY 618 Val-----LeuMetPheIleIleIleSerLysArgGlnLeuAsnAlaTyr 632  
Db 1698 GTCCAGCAGACACCCAGATAGTTGTCTGTCTGTCTCAAGTAATATCGGAAGCAATAATAC 1757  
QY 633 GlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSer 652  
Db 1758 GATGCTATTAAATAATACCTGTGTACAGATTCCTCCATCCCAAGTCAGTGTGTGGGCC 1817  
QY 653 GluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePhe 672  
Db 1818 CGAACCTTTAGGCAAA-----CAGCAAACTGTCTATGGCCATTGCT 1856  
QY 673 TyrGlnIleAlaLeuLysIleAsnAlaLysLeuGlyIleAsnGlnGluLeuAspTyr 692  
Db 1857 ACAAGATTGCCCTACAGATGAACCTGCAAGATGGGAGGA-----GAGCTC---TGG 1904  
QY 693 SerGluIleAlaGluIleSerProGluGluLysGluArgArgLysThrMetPro----- 710  
Db 1905 -----AGGTTGGACATCCCTCCCTGAAG 1925  
QY 711 LeuThrMetTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyr 730  
Db 1926 CTCGTGATGATCGTTGGCATCGATTGTACCATGACATGCAGACTGGG-----CGAGG 1979  
QY 731 SerIleAlaAlaValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIle 750  
Db 1980 TCAATCGCAGGAGTTTGTGCCAGCATCAATGAAGGATGACCCGCTGGTTCACCGCTGC 2039  
QY 751 ValThrGlnGluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThr 770  
Db 2040 ATATTTTCAGAT-----AGAGGACAGAGCTGGTA 2069  
QY 771 AspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsn 790  
Db 2070 GATGGCTCAAAGTCTGCTCGCAAGCGCTCTGAGGGCTTGGATAGCTAGCTAATGAG--- 2126  
QY 791 ArgAlaProAlaHisIleValTyrArgAspGlyValSerAspSerGluMetLeuArg 810  
Db 2127 TACATGCCAGCCGCGATCATCGTTACCGCGATGCGTAGGAGACGCCGCTGAAAAACA 2186



```
QY 811 ValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSer----- 828
Db 2187 CTG-----GTGAACCTACCAAGTGCACACAGTTTGGATTGTCTA 2225
QY 829 GluArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHis 848
Db 2226 AATCCATTGGTAGAGGTTACCAACCTAGACTACGGTAAATGGTGGTGAAGAAAGAGTG 2285
QY 849 AsnThrArgLeuLeuArgMetGluLysAspLysProValValAsnLysAspLeuThr 868
Db 2286 AACACCAGATTTTT----- 2300
QY 869 ProAlaGluThrAspValAlaValAlaValLysGlnTrpGluGluAspMetLysGlu 888
Db 2300 ----- 2300
QY 889 SerLysGluThrGly-----IleValAsnProSerSerGlyThrThrValAspLysLeu 906
Db 2301 ---GCTCAGCTCGAGGAGACTTCAGATCCACTTCTCGACAGTATTATTGATGAG 2357
QY 907 IleValSerLysTyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGlyThr 926
Db 2358 GTTACCAGACCAGAAATGGTATGACTTTTATCTGTGAGCCAGGCTGTGAGAGTGTAGT 2417
QY 927 SerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluVal 946
Db 2418 GTTCTCCACACATTATCAATGTCTATGACACACAGCGGCTGGAAGCCAGACCACATA 2477
QY 947 TyrLysMetThrTyrGlyLeuAlaPheLeu 956
Db 2478 CAGGCTTGACCTACCAAGCTGTGCCATC 2507
```

## RESULT 11

```
US-10-104-047-1676
; Sequence 1676, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1676
; LENGTH: 3138
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1676
```

```
Alignment Scores:
Pred. No.: 2,54e-21 Length: 3138
Score: 278.50 Matches: 189
Percent Similarity: 33.4% Conservative: 144
Best Local Similarity: 19.0% Mismatches: 329
Query Match: 5.2% Indels: 335
DB: Gaps: 40
```

US-10-645-746-3 (1-1020) x US-10-104-047-1676 (1-3138)

```
QY 40 ValLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyr 59
Db 467 GTGAAATGGTTACAAACCTCTTTAACTTAGATTTTCCC---CAAGACTGGCAGCTATAC 523
QY 60 GluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPhePro 79
Db 524 CAGTACCATGTGATATATATTCAGATTAGCATCTAGAGAGCTG----- 568
QY 80 LysLysThrGluIleProLeuProAspArgAlaLysLeuPheTrpGlnHis-----Leu 97
Db 569 -----AGNATGGCTTTACTTTATAGTCATAGTGAACCT 601
```

```
QY 98 ArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAsp 117
Db 602 TCCAAACGAAGCAAAAGCATTCGACGGTCCATC----- 634
QY 118 ThrValTyrSerValCysArgLysLeuAsnThrValThrSerLysMetLeuValSerGluLys 137
Db 635 -----CTTTTCTCTGCACAAAG 652
QY 138 ValValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeuTyr 157
Db 653 CTAGAAGAAAGGTCCAGAGTTGTCAGTGAAACTCAAGAGGTGAGACTATTAAGATG 712
QY 158 ThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGlu 177
Db 713 ACTATCACCCCTG-----AAGAGGAGCTGCCATCA 742
QY 178 LysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysVal 197
Db 743 AGTTCTCCCGTGTGCATCCAGGCTCTC-----AATATCATCTTCAGAAAGATC 790
QY 198 ArgTyrAlaProPheValAsnGluLysValGlnPheAlaLysAsnPheValTyr 217
Db 791 -----CTCAAAAAGTTGTCATGTACCAATTTGGACCGAACTTC---TAT 832
QY 218 AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln 237
Db 833 AATCCTTCAGACCAATGGAAATTTCCCCAGCAC----- 865
QY 238 SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyIleLysGlu 257
Db 866 -----AAATTTATCCCTTTGGCTGGTTTGCATTTCTGTG---TCA 904
QY 258 LeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAla 277
Db 905 TATTTTGAAGGAAGCTCCCTGTTAGTGTGATGTGAGTTACAAAGTCTCTGTTCCACCAG 964
QY 278 ProLysMetSerLeuLeuAspTyrLeuLeu-----LeuIleValAspProGln 293
Db 965 ACGGTTCTGGAAATTCATGACTGCTCTGTCAAAGAACTGCTGTCTGTTCCACCAG 1024
QY 294 SerCysAsn----- 296
Db 1025 ACGTGTGAGAAGCAGCTAATAGGGCTCATTTGCTCTTCAAGATACATAAAGAACCTAC 1084
QY 297 -----AspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLys----- 311
Db 1085 TCCATTGATGACATTGACTGCTCAGTGAAAGCCACACACACACCTTTTCAGAACGGGATGGC 1144
QY 312 -----MetThrIleArgGlnAla 317
Db 1145 ACCGAGATCACCTATGTGGATTACTACAAGCAGCAGTATGATATATTCTGATCGACCTG 1204
QY 318 AlaArgProArgIleArgGlnLeuLeuGluAsnLysLysCysAlaGluValTrp 337
Db 1205 AATCAGCCCCATGCTTGTAGTCTGTTAAAGAAAGAGAGA----- 1243
QY 338 AspAsnGluMetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGlu 357
Db 1244 ---AATGACACAGTCAAGCTCAGCTCGGCCACTG---ATACCTGAGCTCTGC----- 1291
QY 358 AsnSerLeuValTyrLysValThrGlyLysSerAspArgGlyArgGlnAlaLysLysTyr 377
Db 1292 -----TTTCTAACAGGGCTGACTGACCAGGCAACATCT----- 1324
QY 378 AspThrThrLeuPheLysIleTyrGluGluAsnLysPheIleGluPheProHisLeu 397
Db 1325 GATTTCCAGCTGATGAAGGCTGTGGCTGAAAGACACAGT----- 1363
QY 398 ProLeuValLysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGlu 417
Db 1364 -----CTCAGT 1369
```

```
QY 418 ValHisGluLysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPhe 437
Db 1370 CCTTCAGCGGAGCAGCGCTGGCCAGCGTTGTGGCAACATCCAGAGGAATACCAAT 1429
QY 438 LeuLysArgAlaThrArgLysProHisAspTyrLysGluAsnThrLeuLysMetLeuLys 457
Db 1430 GCTCGC----- 1435
QY 458 GluLeuAspPheSerSerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLys 477
Db 1436 ----TTTGAACACTAGAGACTGGGAGCTGCATTTTGGAAAGCCAGATATCTCTGCTGCCGG 1492
QY 478 LeuGlnMetIleGluCysProGlyLysValLeuLysGluProMetLeuValAsnSerVal 497
Db 1493 ATTGCCCTTCAGAA-----AAATATTAAATGCAA-----GACCCACATA 1531
QY 498 AsnGluGlnIleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnVal 517
Db 1532 TGTCAACCTGTGTCTGCTCTGA-CTGGTCCAAGGATATTGCAACTTGCAA----- 1581
QY 518 ValProGluLysGluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnPro 537
Db 1582 -----GATTTTAAATGCACAGCTTT----- 1602
QY 538 CysLeuGluGluAsnAspValValLysPhe----- 547
Db 1603 -----GATACCTGGTTGATTTTATGTAGCGCAGAACTGAATATGTGCCGA 1650
QY 548 ---TyrThrGluLeuIle---GlyGlyCysLysPheArgGlyIleArgIleGlyAlaAsn 565
Db 1651 GAGCTTTCGAACTGCTTCGAGAAGAGTTCAGGTTCCATGGGATTTAATGTGGACTACCC 1710
QY 566 GluAsnArgGlyAlaGlnSerIleMetTyrAspAlaThrLys-AsnGluTyrAlaPheTy 585
Db 1711 CAAAT-----CATAAAGGTACAAGAAATCCAGCTGCATTTGT 1749
QY 585 rLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAlaAlaThrGluAlaL 605
Db 1750 TAGA----- 1753
QY 605 sAsnMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPheIleIleLe 625
Db 1754 ----GCTATACAGCAATATGTTGATCCTCATGTTCCAGTGGTAAATGTGCATTCTGCCTTC 1809
QY 625 rLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAspHisThrIleGlyVa 645
Db 1810 T---AATCAGAGACCTATTATGATTCCTAATAAAATATTTGAGCTCAGACTGCCCAT 1866
QY 645 lAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeuArgHisGl 665
Db 1867 CCCAAGCCAAATGTGCTGTGCTCGGACCTTGAATAA-----CA 1905
QY 665 uLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsnAlaLysLeuGlyGl 685
Db 1906 CGGCATGATGATGATATCCGCCAACAGATCGTATGCAGATGACTTGCAGCTCGGAGG 1965
QY 685 yLeuAsnGlnGluLeuAspTyrSerGluIleAlaGluIleSerProGluGluLysGluAr 705
Db 1966 C-----GAGCTG---TGCGCT-----GTGGAATA----- 1987
QY 705 gArgLysThrMetProLeu-----ThrMetTyrValGlyIleAspValThrHisProth 723
Db 1988 -----CCTTTAAAGTCCCTGATGGTGGTATGATGATGATGATGATGATGATGATGATG 2034
QY 723 rSerTyrSerGlyIleAspTyrSerIleAlaValValAlaSerIleAsnProGlyGl 743
Db 2035 ACTCAGCAAG-----GAGCTGATGTTGTTGATGCTGCGTGGCCAGTGTAAACCCAGAA 2088
QY 743 yThrIleTyrArgAsnMetIleValThrGlnGluGluCysArgProGlyGluArgAlaVa 763
Db 2089 CACCAAGGTGGTTTCCCGCTGTATCCTTCAG----- 2119
QY 763 lAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheVal-----Ly 779
```

```
Db 2120 -----AGAACAAATGACTGATGTTCCAGATTGCTTTGAAGTTTTCATGACTGGAGC 2169
QY 779 sLeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAlaHisIleValValTy 799
Db 2170 ACTCAACAAATGGTACAAAGTACAATCATGAT-----TTGCCAGCAGCGATAATTGTGTA 2223
QY 799 rArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLe 819
Db 2224 CGGTCTGCTGTAGGGATGGTCAGCTGGAAACACTTTATTGAATATGAAGTCCACAGCT 2283
QY 819 uLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLysTy 839
Db 2284 GCTGAGCAGTGTGGCAGAA-----TCCGCTCAATATACCAGCTCAAGACT 2328
QY 839 rThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgArgMetGluLysAs 859
Db 2329 GTCCGTGATTGTGTGTCAGAGAGAGTGCATGCCAGGATTTCTTTACCGAAATGAACGCA 2388
QY 859 pLysProValValAsnLysAspLeuThrProAlaGluThrAspValAlaAlaAlaVa 879
Db 2389 T----- 2389
QY 879 lLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyIleValAsnProSerSe 899
Db 2390 -----GTACAGAACCCCCCACT 2406
QY 899 rGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPhePheLeuAlaSe 919
Db 2407 TGGCACTGTGTGGATTTTCAGAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAG 2466
QY 919 rHisHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLy 939
Db 2467 CAGGTGGCTGCCGGGGAACCTGTAGTCTACCTACTATAATGATCATCTATGATGACAA 2526
QY 939 sGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaAr 959
Db 2527 CGGCTTGAAGCCCGACCATATGCAGAGACTTACATTCAAATTTGCGCACCTGTACTACAA 2586
QY 959 gCysArgLysProIleSerLeuProValHisTyrAlaHis 974
Db 2587 CTGGCCGGGCATAGTCAGTGTGCCAGCACCATGTCAGTATGCTCAC 2632
```

## RESULT 12

US-09-248-796A-6819  
; Sequence 6819, Application US/09248796A  
; Patent No. 6747137

## GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 6819

; LENGTH: 771

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-248-796A-6819

## Alignment Scores:

Pred. No.:	8,366-21	Length:	771
Score:	265.00	Matches:	75
Percent Similarity:	44.6%	Conservative:	54
Best Local Similarity:	26.0%	Mismatches:	112
Query Match:	5.0%	Indels:	48
DB:	3	Gaps:	9

US-10-645-746-3 (1-1020) x US-09-248-796A-6819 (1-771)

```
QY 714 TyrValGlyLeuAspValThrHisProThrSerTyrSerGlyLeuAspTyrSerIleAla 733
Db 16 TACTTGGTCTGTGATGTTACT-----TCATCACAACAGGAGAAATTTATTCTGAATCA 69
QY 734 AlaValValAlaSerIle---AenProGlyGlyThrIleTyrArgAenMetIleValThr 752
Db 70 GTTCTATTGCTTCTATTGTTGGTAGAGATGGAATTTTCAATAAATTC-----120
QY 753 GlnGluGlyCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIle 772
Db 121 -----CCTGCATCAGTTCTGATTCCAACTGGTGACAAGAGTT--ATT 162
QY 773 LeuGluAlaLysPheValLysLeuLeuArg---GluPheAlaGluAenAenAspAenArg 791
Db 163 GCTGATGTTAAAGTAGTATGTTTGGAAACGATTAGAAAAATTTCCATAAAAAATTTGGGAAA 222
QY 792 AlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgVal 811
Db 223 TTACCTAGTAAAGTATTATTATTCGTGATGGAGTCTCTGAAGGACAATATACACCCATA 282
QY 812 SerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAsp 831
Db 283 TTAAGAAGAATTAACAAAAATTAAGCCGCTTTTAATGAATATGTAATTAATGAAGAAC 342
QY 832 GlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHisAenThrArg 851
Db 343 ATCCCTAAATATCTCCAAACCAATACATTTATGTTGTTAAAGACATCATACAAGA 402
QY 852 LeuLeuArgArgMetGluLysAspLysProValValAenLysAspLeuThrProAlaGlu 871
Db 403 TTTATT-----CCTATTCTCATGATAAT-----423
QY 872 ThrAspValAlaValAlaValLysGlnTrpGluGluAspMetLysGluSerLysGlu 891
Db 424 -----GCCGATCATCCAAAAACCAAAAAACAA 450
QY 892 ThrGlyIleVal-----AenProSerSerGlyThrThrValAspLysLeuIleVal 908
Db 451 ATAGCTGTCACCAAGTAATGAAATATGTCATGCTGTACAAACCGTTGATCGAGAAATCACT 510
QY 909 SerLysTyrLysPheAspPheLeuAlaSerHisIleGlyValLeuGlyThrSerArg 928
Db 511 TCTCCAGCATTTTTCATTTTACGTTCAATCTCAACAATCATTAACAGGTACTGGAATA 570
QY 929 ProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLys 948
Db 571 CCAGCTCATATTATGTTTACATGATGAAATAATATATCTTCAGATACTATACAAAAA 630
QY 949 MetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProLysSerLeuProVal 968
Db 631 ATCACTTATGATTATGTCATCTATTTTAGTAGAGTACTATAATCAGTCAAAAGTTGTCGCC 690
QY 969 ProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLys 988
Db 691 GCAGCTTATATGCCGATTATTATGTTACTAGAGGTAGAGAT-----732
QY 989 GluHisTyrIleGlyAspTyrAlaGln 997
Db 733 -----TATATTATGTTTGTCTAAA 753
```

## RESULT 13

```
US-09-297-648-2528
; Sequence 2528, Application US/09297648
; Patent No. 6964868
```

## GENERAL INFORMATION:

```
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
```

```
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassen, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, William Lee
; APPLICANT: Stache-Crain, Birjit
; TITLE OF INVENTION: No. 6964868el Human Genes and Gene Expression
; TITLE OF INVENTION: Products II
; FILE REFERENCE: 2300-1481
; CURRENT APPLICATION NUMBER: US/09/297,648
; CURRENT FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,666
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/080,515
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 5252
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2528
; LENGTH: 734
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(734)
; OTHER INFORMATION: n = A, T, C or G
US-09-297-648-2528
```

## Alignment Scores:

Pred. No.:	7,61e-20	Length:	734
Score:	256.50	Matches:	57
Percent Similarity:	50.6%	Conservative:	23
Best Local Similarity:	36.1%	Mismatches:	47
Query Match:	4.8%	Indels:	31
DB:	4	Gaps:	3

US-10-645-746-3 (1-1020) x US-09-297-648-2528 (1-734)

```
QY 833 GluAspProGluProLysTyrThrPheIleValIleGlnLysArgHisAenThrArgLeu 852
Db 118 AAAGACTATCAACCTGGAATACCTACATTGTTAGTTCAGAGAGACATCACACTCGATTA 177
QY 853 LeuArgArgMetGluLysAspLysProValValAenLysAspLeuThrProAlaGluThr 872
Db 178 TTT---TGTCGTGATAGGACAGAAAGGGTTGGAAGAAGTGGCAATATCCAGCT-----228
QY 873 AspValAlaValAlaAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThr 892
Db 228 -----228
QY 893 GlyIleValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLys 912
Db 229 -----GGAACAACAGTTGATACAGACATTACACACCCCATATGAG 267
QY 913 PheAspPhePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyr 932
```

```

Db      268 TTGATTTTACCTCTGTAGCATGCTGGNATACAGGTTACCATCGTCTCTTCACACTAT 327
Qy      933 ThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGly 952
Db      328 CATGTTTATGGGATGATAACTGCTTTACTGCAGATGAACCTTCAGCTGCTAACTTACCAG 387
Qy      953 LeuAlaPheLeuSerAlaArgCysArgLysPheProIleSerLeuProValProValHisTyr 972
Db      388 CTCCTGCCACACTTACGTACGCTGTACACGATCTGTTTCTATACCTGCACACGCGTATTAT 447
Qy      973 AlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990
Db      448 GCTCACTGTGACATTAGACCCAGA---TATCACTTGTGGACAAAGAACAT 498

RESULT 14
US-09-873-737A-1
; Sequence 1, Application US/09873737A
; Patent No. 6723534
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; FILE REFERENCE: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; CURRENT APPLICATION NUMBER: US/09/873, 737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3047
; TYPE: DNA
; ORGANISM: Drosophila sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84)..(2612)
; NAME/KEY: misc feature
; LOCATION: (120)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
; NAME/KEY: misc feature
; LOCATION: (399)
; OTHER INFORMATION: n=a or t, Xaa=Leu or Ile
; NAME/KEY: misc feature
; LOCATION: (2436)
; OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
US-09-873-737A-1

Alignment Scores:
Pred. No.: 9,96e-13 Length: 3047
Score: 205.00 Matches: 175
Percent Similarity: 33.7% Conservative: 137
Best Local Similarity: 18.9% Mismatches: 270
Query Match: 3.8% Indels: 346
DB: 3 Gaps: 45

US-10-645-746-3 (1-1020) x US-09-873-737A-1 (1-3047)

Qy      231 AspProAsnArgPheGluGlnSerLeuGluValAlaProArgIleGluAla----- 247
Db      225 GAGGCTTCAAGAGAGAGAGAGAGCTCTCGAGGAAGCTCCCGAGGCTGAGCGTGGCGCGCA 284
Qy      248 -----TrpPheGlyIleTyr-----lleGly 254
Db      285 GAGCGAAAGCCGTGGGGTGCACCAATATGATTACCTGAATACCGTCCGCTTGAGCTGGTA 344
Qy      255 IleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPhe 274
Db      345 TCCAGAAGGAAGCAACCATGCGTCCCGGTCTATG-----CTGCAGACGAACTTT 392
Qy      275 TyrAsnAla-----ProLysMetSerLeuLeuAspTyrLeuLeuIleValAsp 291

```

```

Db      393 TTTTCANTAACCAACGACCGGAATGCGGATCGTTTATTATCAGCTGGAGTTGTGCGCG 452
Qy      292 ProGlnSerCysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLys 311
Db      452 ----- 452
Qy      312 MetThrIleArgGlnAlaAlaArgProArgIleArgGln----- 324
Db      453 ---ACCATCGAG-----AATCCTCGTGTCCGTATGGGAGTTTGTGCCAATCANGCT 500
Qy      325 -----LeuLeuGluAsnLeuLysLeuLysCysAlaGluValTyr 337
Db      501 AACCTTCTGGGATCAGGCTATCTATTCGACGACTGCACACTGTTCCACCACGAGAAATTC 560
Qy      338 AspAsnGluMetSer-----ArgLeuThr 345
Db      561 GAGCAGGAAATCACGCTGCTCAGCGGAAAGTCGAAAGTCGACATTCGAATACAAGATATCC 620
Qy      346 GluArgHisLeuThrPheLeuAspLysCysGluGluAsnSerLeu-----ValTyrLys 363
Db      621 ATAAAGTTTCGTTGGATTTCATATCGTGTCTGAGCCCGCTTTTTCGAAGTCTTAAATCTA 680
Qy      364 ValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLys 383
Db      681 ATATTGCGCGCTCGATGAGGCGCTAAATTGGGAATTAGTTGGCGCTAATCTCTTTGAT 740
Qy      384 IleTyrGluGluAsnLysPheIleGluPheProHisLeuProLeuValLysValLys 403
Db      741 -----CCCGAGCTAAGATCGAA 758
Qy      404 SerGlyAlaLysGluTyrAlaValProMet-----Glu 414
Db      759 -----ATAAGGAGGTTCAAAATGAGCTATGCGCGGCTATGAGACATCGATTCGTCTAG 812
Qy      415 His-----LeuGluValHisGluLysProGlnArgTyrLysAsn 427
Db      813 CACGAAAGATATTTATTGGGCACCGAAATACTCACAAAGTTATGCGCAGACGAG 872
Qy      428 ArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAsp 447
Db      873 ATCTACGACATA-----ATGCGACGTTGCTCACACAAATCGGCTCGT 914
Qy      448 TyrLysGlu-----AsnThrLeuLysMetLeuLysGluLeuAspPheSerSer 463
Db      915 CATCAGGACGAAGTACGGGTAAATGTTTGGACTTGATTGCTCTTACGGATTACAATAAC 974
Qy      464 Glu-----GluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIle 481
Db      975 AGAACTTATCGTATCAATGATGTCGAC---TTTGGACAAACTCCGAAA---TCAACATTC 1028
Qy      482 GluCysProGlyLysValLeu----- 488
Db      1029 AGTTGCAAGGCTAGAGATATCATGTTTCGTGGAATACTATCTCACTAAATATAATATACGC 1088
Qy      489 -----LysGluProMetLeuValAsn----- 495
Db      1089 ATTCGCGACCAATACGCGCTGCTGATTTCACAAAATAGGACACAGGCTCTAAAAAAT 1148
Qy      495 ----- 495
Db      1149 AACGCTAGCGAATTAGTGTACTAATTCTCTGAGCTCTGCCGAGTGACTGGGCTCAATGCC 1208
Qy      496 -----SerValAsnGluGlnIleLysMetThrPro 505
Db      1209 GAGATGCGCTCAAACTTTTCAGCTTATGCTGTCATGAGCAGTTATACGCAATGAACCC 1268
Qy      506 Val-----IleArgGlyPheGlnGluLysGlnLeuAsnValValProGlu 520
Db      1269 AAACACGCACTGATCGATTGGCGCTTTTAAACCAACCGTTTACAAAAC---ACTCCAGAA 1325
Qy      521 LysGluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGlu 540

```



```
QY 782 gGluPheAlaGluAsnAsnArgAlaProAlaHisIleValValTyrArgAspG1 802
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1960 CCAATAT---CAACATGAGCATAGGAAGCTGCCATCTCGAATCGTATTTATCGAGACGG 1904
QY 802 yValSerAspSerGluMetLeuArgValSerHis-----AspGluLeuArgSe 818
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1903 CGTAGCTCCGGCTCTCTAAGCAGCTTTTGAATTTGAAGTCAAGGACATCATTCAGAA 1844
QY 818 rLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLy 838
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1843 GTTGAATACTGAATCCGCCGCTCCAGCTAAGC-----CCACCGCA 1802
QY 838 sTyrThrPheIleValIleGlnLysArgHisenThrArgLeuLeuArgArgMetGluLy 858
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1801 ATTAGCTTATATTGTGTAAACAGATCCATGAACACGCGCTTCTCCTCAACGGACAA-- 1744
QY 858 sAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAlaValAlaAl 878
Db ----- 1744
QY 878 aValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyIleValAsnProSe 898
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1743 -----AATCCTCC 1736
QY 898 rSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPheLeuAl 918
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1735 GCCTGTACTATATTGTATGAGCTTATAACTCTGCCCGAGAGATAGACTTTTATCTGTT 1676
QY 918 aSerHisHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAs 938
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1675 CTCGCAACAAGTTCGTACGGGTACAGTGTCCCGACCACTACAAATGTTCTTTATAGCAG 1616
QY 938 pLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeu----- 956
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1615 CATGGTCTCTCACCGGAGAAAATGCAAAACCTTACGTACAAAGATGTGCCACTTGTACTA 1556
QY 957 -----SerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHi 974
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1555 CAATTGGTCCGGCACCACACGAGTGCCA----- 1528
QY 974 sLeuSerCysGluLysAlaLysGluLeu 983
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1527 -GCCGTTTCCAGTACGCTAAGAAGTTA 1501
```

Search completed: July 5, 2006, 23:28:20  
Job time : 471 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

# QM protein - protein search, using sw model

Run on: July 5, 2006, 13:46:42 ; Search time 23 Seconds  
(without alignments)  
1189.943 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 5349  
Sequence: 1 MASNPELEKGFVHSLDPE.....RHEMHFLQTNVYKPGMSFA 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	577	10.8	857	US-11-192-437-1	Sequence 1, Appli
2	566	10.6	859	US-11-192-437-2	Sequence 2, Appli
3	555.5	10.4	860	US-11-192-437-3	Sequence 3, Appli
4	555	10.4	861	US-11-192-437-4	Sequence 4, Appli
5	535	10.0	1050	US-10-953-349-4064	Sequence 4064, Ap
6	531	9.9	870	US-10-953-349-4065	Sequence 4065, Ap
7	509	9.5	1052	US-10-449-902-53188	Sequence 53188, A
8	507.5	9.5	822	US-10-953-349-4086	Sequence 4086, Ap
9	500	9.3	1011	US-10-449-902-56159	Sequence 56159, A
10	431.5	8.1	876	US-10-449-902-37356	Sequence 37356, A
11	427	8.0	1145	US-10-997-086-5	Sequence 5, Appli
12	421.5	7.9	517	US-10-449-902-38630	Sequence 38630, A
13	403.5	7.5	457	US-10-953-349-18475	Sequence 18475, A
14	403.5	7.5	458	US-10-953-349-18474	Sequence 18474, A
15	403.5	7.5	501	US-10-953-349-18473	Sequence 18473, A
16	396.5	7.4	527	US-10-953-349-19398	Sequence 19398, A
17	370.5	6.9	277	US-11-192-437-40	Sequence 40, Appli
18	370.5	6.9	488	US-10-953-349-19399	Sequence 19399, A
19	366	6.8	630	US-10-449-902-47718	Sequence 47718, A
20	361.5	6.8	471	US-10-953-349-19400	Sequence 19400, A
21	352.5	6.6	277	US-11-192-437-41	Sequence 41, Appli
22	344.5	6.4	277	US-11-192-437-39	Sequence 39, Appli
23	341.5	6.4	287	US-11-192-437-42	Sequence 42, Appli
24	302	5.6	292	US-10-449-902-54734	Sequence 54734, A
25	284	5.3	266	US-10-449-902-31632	Sequence 31632, A

26	284	5.3	286	6	US-10-449-902-32271	Sequence 32271, A
27	260.5	4.9	666	7	US-11-293-697-3968	Sequence 3968, Ap
28	234.5	4.4	763	6	US-10-449-902-43163	Sequence 43163, A
29	156	2.9	3113	6	US-10-505-928-325	Sequence 325, App
30	151	2.8	1328	6	US-10-504-973-32	Sequence 32, Appli
31	143.5	2.7	1502	6	US-10-471-571A-1696	Sequence 1696, Ap
32	142	2.7	956	7	US-11-293-697-3037	Sequence 3037, Ap
33	140	2.6	2871	6	US-10-505-928-100	Sequence 100, App
34	139	2.6	1354	7	US-11-270-653-1	Sequence 1, Appli
35	138	2.6	1674	6	US-10-511-937-2587	Sequence 2587, Ap
36	136	2.5	108	6	US-10-449-902-32719	Sequence 32719, A
37	136	2.5	436	6	US-10-471-571A-5206	Sequence 5206, Ap
38	133.5	2.5	1116	6	US-10-449-902-47106	Sequence 47106, A
39	129	2.4	824	7	US-11-293-697-4573	Sequence 4573, Ap
40	127.5	2.4	1003	6	US-10-471-571A-4826	Sequence 4826, Ap
41	126.5	2.4	1517	7	US-11-257-500-9	Sequence 9, Appli
42	126.5	2.4	1526	7	US-11-257-500-15	Sequence 15, Appli
43	125.5	2.3	972	6	US-10-480-962-4	Sequence 4, Appli
44	124.5	2.3	94	7	US-11-192-437-13	Sequence 13, Appli
45	124.5	2.3	94	7	US-11-192-437-14	Sequence 14, Appli

## ALIGNMENTS

### RESULT 1

US-11-192-437-1  
; Sequence 1, Application US/11192437  
; Publication No. US20060141600A1  
; GENERAL INFORMATION:  
; APPLICANT: Joshua-Tor, Leemor  
; APPLICANT: Song, Ji-Joon  
; APPLICANT: Hannan, Gregory J.  
; APPLICANT: Liu, Jidong  
; APPLICANT: Carmell, Michelle A.  
; APPLICANT: Rivas, Fabiola  
; APPLICANT: Marsden, Carolyn  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO  
; FILE REFERENCE: ARGONAUTE PROTEINS  
; CURRENT APPLICATION NUMBER: US/11/192,437  
; CURRENT FILING DATE: 2005-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,297  
; PRIOR FILING DATE: 2004-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,269  
; PRIOR FILING DATE: 2004-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 857  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-192-437-1

Query Match	10.8%	Score 577;	DB 7;	Length 857;
Best Local Similarity	23.0%;	Pred. No. 3e-28;		
Matches 227;	Conservative 171;	Mismatches 365;	Indels 222;	Gaps 37;
QY	38	KKVLLLVNFKSSKIYDREYEVKMTKEVLNRPKPGPKKTEIPIDRAKLFQHL	97	
DB	34	KPKLLANFYVDIPKIDVHYEVDIKPKD-----CPRVRREVVE-----YMV	77	
QY	98	RHEKKQTDPILEDY--VFDEKDTVSVCRINTVTSKMLVSEKV---VKQSEKKDKDLE	152	
DB	78	QHFKPQ---IFGDRKPVYDGKNIY-----TVALPIGNERVDFVTIPGEGKD-RIFK	127	
QY	153	KKILYTMILTYKKFHLNFSRENPEKDEANRSYKFLKNVMTOKVYAPVFNVEIKVQPA	212	
DB	128	VS1KKWLAIYSWRMLHEALVSGQIPVPLESVQALDVAHRHLAS--MRYTPV-----	175	
QY	213	KNFVVDNNSILRVPSFHDNPFQESLEVAPEATMFGIYIGIKELFDGEPVLNFAVDK	272	
DB	176	-----GRSFFSPPEGYHP-----LGGGREGVWFGHQSVRPAM-WKQMLNIDVSAT	220	



QY 877 AAVQWEDMKESKETGIVNPSSTGTTVDKLVSKYKEDFFFLASHGVLGTSRPHGYTMY 936  
Db 733 -----GTTVDTKITHPTFDFYLCSHAGIQTSPRSHYHLW 769  
QY 937 DDKGMSQDEVVYKMTYGLAFLSARCKPISLPVPVHYAHLSCAKELYRTYKEH 990  
Db 770 DNRFSDELQILTYQLCHTVTRSVSIPAPAYAHVAFRAR-YHLVDKEH 822

RESULT 3  
US-11-192-437-3  
; Sequence 3, Application US/11192437  
; Publication No. US20060141600A1  
; GENERAL INFORMATION:  
; APPLICANT: Joshua-Tor, Leemor  
; APPLICANT: Song, Ji-Joon  
; APPLICANT: Hannon, Gregory J.  
; APPLICANT: Liu, Jidong  
; APPLICANT: Carmell, Michelle A.  
; APPLICANT: Rivas, Fabiola  
; APPLICANT: Marsden, Carolyn  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO  
; TITLE OF INVENTION: ARGONAUTE PROTEINS  
; FILE REFERENCE: CSHL-P01-014  
; CURRENT APPLICATION NUMBER: US/11/192,437  
; CURRENT FILING DATE: 2005-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,297  
; PRIOR FILING DATE: 2004-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,269  
; PRIOR FILING DATE: 2004-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 860  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-192-437-3

Query Match 10.4%; Score 555.5; DB 7; Length 860;

Best Local Similarity 23.2%; Pred. No. 6.5e-27;

Matches 230; Conservative 149; Mismatches 377; Indels 237; Gaps 35;

QY 38 KKVLLVNNFKFSKLYDRYYEYVQWTKVNLNRPKPPKTEIPIDRAKLFQWHL 97  
Db 28 KPICKLANCFQVBPIDVLYEVDIKPKD-----CPRRVREVVDSD---MVQHP 74  
QY 98 RHEKKQTDFTLEDYVFEKDTVYVSVCLNTVTSKMLVSEKVKVSKDSEKKDKLEKKILY 157  
Db 75 ----KVTIFGDRPVDGKSLTANPLPVATTG-----VLDVTLPGEKGKDRPFKV 123  
QY 158 TMLTYRKXPHLNFNSRENPEKBEANRSYKFLKNVMTQKRYAPFVNEEIKVQFAKNFY 217  
Db 124 SIKFVSRVSNHL-----LHEVLGTGLPEPL---ELDKPISNPNVH 161  
QY 218 DNNSILR-----VPSFHDNPRFQSLVAPRIEAWFGIYIGIKBLFDG 261  
Db 162 AVDLVRLHPSMKYTPVGRSFFSAPGEGYDHP-----LGGREVMFQGHQSVRAM-W 212  
QY 262 EPLVNEAIVDKLFYNAPKSNLLDYLLIVD-----PQSCNDVDRKDLTKLMACKMTI 314  
Db 213 KMLNLDVSNATFYKA--QPVIOFMCVELDIHNDIQPRPLTDSHRVKFTKEIKGLKVEV 270  
QY 315 QMARPRIRQLLENLKLKCAEVWDMNSRLTERHLTF-LDLCEBSLVYKVTGKSDRGRN 373  
Db 271 THCGTWRKRYVCN-----VTRRPASHQTFPLQL-----EN 301  
QY 374 AKKYDTTLFKIYEENKFP-IEFPHLPLVKKVSAKAYVAPMEHLEVHEKPPQRYKNRDLV 432  
Db 302 GQTVERTVAYQFREKTYLQKYPHLFCLQVGQKQKHTYLPLEVNT-VAGQRCIKLTON 360  
QY 433 MQDKFLKRAIRKPHDVKENTKMLKELDFSESSEINLNFVERFGLCSKLQMIETCPKVLKEP 492

Db 361 QTSMIKATARSAPDROBBISRLVRSANYETDP--FVQEFQKVRDEMAHVTGRVLPAPM 418  
QY 493 LVNSVNEQIKMTVP-----IRGFOEKQLNV-VPEKELCCAVFVNVNAGNCLBENDVVK 546  
Db 419 LQVGRNRRTVATPSHGVDMDRG---KQFHTGVEIKWALACFATQ-----RQREE--ILK 469  
QY 547 FYTELI-----GGCKFRGIRIGANENRGAQSIYDATQNEYAFYKNCNTLNTGIRPEIA 600  
Db 470 GFTDQLRKISKDAGMPTQGFPCFYAQAGADSV----- 502  
QY 601 ATEAKMFERLPDKKEQKVLFIISKQNLNAYGVPVHYCDHTTIGVANOHITSETVTKALA 660  
Db 503 ----EPNFRHLKNTYSLGLIILVILPKPTPVYAEVRVGVDTLLGNATQCQVQKVIKT-- 556  
QY 661 SLRHEKSGKRIFYQIALKINAKLGGINOELDSWSEIAEISPEEKERRRKTMTPLMTYVGIDVT 720  
Db 557 -----SPQTLNLCILKINVLGGINNIL-----VPHQPSVFPQPV-IFLGADVT 600  
QY 721 HPTSYSGIDYSIAAVVASINPGGTIYRMVITVQEECPGRBAVANGRERTDILB--AKFV 778  
Db 601 HPPAGDGKPKSIAAVVGSMDAHPFSRYCATVRVQ-----RPROBIIQDLASWV 647  
QY 779 K-LLRBEAENNDRAPAHIVVYRDGVSDEMLRVSHDLSLSEVKQPMSERDGEDEP 837  
Db 648 RELIQPKSTRFK-PTRIIFYRDGVSQGRQVLYELLALREACISL-----EKDYQP 701  
QY 838 KYTFIVIQKRNTRLRLRMKDKPVVNNKOLTPAETDVAAVAVQWEDMKESKETGIVNP 897  
Db 702 GITVIVVQKHHHTLP-CADRTERRVGRSGNIPA----- 733  
QY 898 SSGTTVDKLVSKYKEDFFFLASHGVLGTSRPHGYTMYDDKGMQDEVVYKMTYGLAFLS 957  
Db 734 --GTTVDTDITHPYBDFYLCSHAGIQTSPRSHYHLVMDNCFNCTADELQILTYQLCHTY 791  
QY 958 ARCKKPISLPVPVHYAHLSCAKELYRTYKEH 990  
Db 792 VRCRVSIPAPAYAHVAFRAR-YHLVDKEH 823

RESULT 4  
US-11-192-437-4  
; Sequence 4, Application US/11192437  
; Publication No. US20060141600A1  
; GENERAL INFORMATION:  
; APPLICANT: Joshua-Tor, Leemor  
; APPLICANT: Song, Ji-Joon  
; APPLICANT: Hannon, Gregory J.  
; APPLICANT: Liu, Jidong  
; APPLICANT: Carmell, Michelle A.  
; APPLICANT: Rivas, Fabiola  
; APPLICANT: Marsden, Carolyn  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO  
; TITLE OF INVENTION: ARGONAUTE PROTEINS  
; FILE REFERENCE: CSHL-P01-014  
; CURRENT APPLICATION NUMBER: US/11/192,437  
; CURRENT FILING DATE: 2005-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,297  
; PRIOR FILING DATE: 2004-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,269  
; PRIOR FILING DATE: 2004-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 861  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-192-437-4

Query Match 10.4%; Score 555; DB 7; Length 861;  
Best Local Similarity 23.0%; Pred. No. 7e-27;  
Matches 227; Conservative 161; Mismatches 378; Indels 220; Gaps 36;  
QY 38 KKVLLVNNFKFSKLYDRYYEYVQWTKVNLNRPKPPKTEIPIDRAKLFQWHL 97





```

QY 81 K---TEIPIDRAKLFQHLRHEKKQTDILEDYVDEKDTVYVSVCLNTVTSKMLVSEK 137
DB TRINRVVMSLARL-----HRESHLGGI--SFAYDGSKALY-----TAGKLFFDS 278
QY 138 VVKDSEKKDEKLEKKILYTMILTYRK-----PHLNFSPRENPEKDEANRSYKPLKNV 192
DB 279 MFKIKLGHLEIEYKV-----TIRAGQADLHLHFEIAGRQDSQ-QOTIQALDVV 331
QY 193 MTQKRYAPFVNEEIKVQPAKNFVYDNNLSILRVPSFHDNRFQESLEVAPRIEAWFIY 252
DB 332 L---RESPSLNVI-----VSRSFYSTMFGRQ--DIGDGLCEWKGY 368
QY 253 IGIKELPDGEPVNFALVDKLFYNAPKMSLLDYL--LIYDPOSCN--DDVRKDKYTKLMAG 310
DB 369 QSLRPTQWGLS--LNIDISPTPFK--PISVVEYVKNCLGTPTNANGPDPRRFLS----- 419
QY 311 KMTIRQAARPRIRQLLENLKLCAEVDNEMSLRTERHLTFLDLCEENSLVYKVTGKSDR 370
DB 420 -----DIDRLKVKKALRGVRETHQGSKYKIT--ITSEPLSQLN---FSMDGTQ- 468
QY 371 GRNAKYDOTTLPKIYENKKF-IEFPHPLPVKVGAKAYAVPMHELVHEKPKQRYKNRI 429
DB 469 -----TVIQFSQRYKRLQVTSWPCLOSGNPSNPIYLPMEVCTIVE--QORYSKKL 518
QY 430 DLVMQDKFLKRATRPHDKYENTLKMKELDFSSBELNVERFGLCSKLOMIECPKVLK 489
DB 519 NDKQVTLGLLRATCQPPKQKEQKIEMVQHNYPADKV--VSDFRINISQMATMPARVLP 576
QY 490 EPLVNSVNEQIKMTVIRGFQEBKQLNVVPEKELCCAVFVNENETAGNCPLEENDVVKFYT 549
DB 577 APTL-----RYHDS-----GKEKTC-----NPRVGQWNNIN--K 603
QY 550 ELIGGC-----KFRGIRIGANENRGAQSIMYDATKNEYAFYKNCNTLNTGIGRFEIAA 601
DB 604 KMGVGAUVQKWCWPSRMHIDA-VHRLCGELVYTCNAIGMFVNEPTEVGSAAPNNIE 662
QY 602 TEAKMFEPLPKEQKQVLMFIIISKRQLNA-YGVKHYCDHTIGVANQHITSETVTKALA 660
DB 663 AALSNIHTRAPQ-----LQLLIVLPDNGYGRIKRVCTELGIVSQCLKP---GRKLL 714
QY 661 SLRHEKSGRIFFQIATKINAKLGINQELDWSAEIAEISPEEKERRKTMF-----LTMV 715
DB 715 SL-----DRQPLENSLKNVAGGNSVL-----QRPLVPGGLENTIIF 755
QY 716 GIDVTHPTSYGIDVSIAAVAVASIN-PGGTIYRNMI-----VTQEEC 756
DB 756 GADVTHPASGEDSSASIAAVASMDWPETIKYKALVSAQPPRQEIIQDLFTMTVAQNAD 815
QY 757 RGERAVAHGRERTDILEAKFVKLLREFAEANNDRAPAHIVYVRDGVSDSEMLRVSHDEL 816
DB 816 APAQK--AEGSKNFCGGMFRELMSFYSKNAKRPQRIIIFRVDGVSQGLHLVLLYEM 873
QY 817 RSLKSEVKQFMSEKDEDP--RPKYTFIVIKRHNTRLRRMEKOKPVVN--KDLTPAET 872
DB 874 DAIKAIASL-----DRAYPLTVFVVQKXHHRLF-----PEVHGQDLT---- 915
QY 873 DVAAVAKQWEEDMKESKETGINPSSGTTVDKLIIVSKYKFOFFFLASHHGVIGTRSPGHY 932
DB 916 -----DRSGNVRP--GTWVDTNCHPSEFDFYLCSHAGIQTSRPTHY 956
QY 933 TVMYDDKMGSDQEVYKMTYGLAFLSARCKPISLPVPHVAHLSCEKAK 981
DB 957 HVLHDENRFSADOLQMLTYNLCTYARCTRSVWPPAYIAHLAARFR 1005

```

## RESULT 8

```

US-10-953-349-4066
; Sequence 4066, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

```

```

; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4066
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-4066

```

```

Query Match          9.5%; Score 507.5; DB 6; Length 822;
Best Local Similarity 22.4%; Pred No. 5,9e-24;
Matches 210; Conservative 159; Mismatches 335; Indels 235; Gaps 36;

QY 113 FDEKDTVYVSVCL--NTVTSKMLVSEKVVVKDSEKDEKDKLEKKILYTMILTYRKFFHLN 170
DB 20 YDGRKSLYTAGPLFPNSKEFRINLLDEEVGAGQRRER---EFKVVIKLV-ARADLHLHG 75
QY 171 ESRENPEKD--EENRSYKF-LKNVMTQKRVYAPFVNEEIKVQPAKNFVYDNNLSILRVPE 227
DB 76 MFLEGKQSDAPOEALQVLVDIVLRELPFGRIYIP-----VGR 112
QY 228 SPHDNRRPEOSLEVAPRIEAWFIYIGIK-----ELFDGEPVNFALVDK 272
DB 113 SFYSPD-TGKQSLGDLGLESWRGFYQSIRPTQMGSLNIDMSSTAFIEANPIQF----- 166
QY 273 LFYNAPKMSLLDYLIIYDPOSCNDDVRKDKLTKLMAGQMTIRQAARPIRQLLENLKLK 332
DB 167 -----VCDLLNRDISRRPLSDADRVKIKKALRGVKEVTHRGNNRRKRYISGL--- 214
QY 333 CAEVDNEMSLRTERHLTF-LDLCEENSLVYKVTGSDRGNKAKYDITLTPKIYENKKF 391
DB 215 -----TAVATRELTFPVD-----ERNTOK---SVVEYFHEYGF 245
QY 392 -IEFPHPLPVKVGAKAYAVPMHELVHEKPKQRYKNRIDLVNQDKFLKRATRKPHDYKE 450
DB 246 RIQHTQLPCLQVGNRNPNYLPMEVKIVE--QORYSKRLNERQITALLKVTQCORPIDREK 304
QY 451 NTLKMLKELDFSSBELNVERFGLCSKLOMIECPKVLKBPML-VNSVNEQIKMTVIRG 509
DB 305 DILQTVQLNDYAKD--NYAQEFGIKISTLASVEARILPPPLWKVHESREGTCLP--- 358
QY 510 FOEKQLNVVPEKELCCAVFVNENETAGN-PCL-----EENDVVKFYTELIGGCKPFGIRI 562
DB 359 -QVGQWNNMKK-----MINGTVNNWJICINFSQVQDNLARTFCQELAQACVYSGM-- 409
QY 563 GANENRGAQSIMYDATKNEYAFYKNCNTLNTGIGRFE-----IAATEAKNMFERLPDKEQKV 618
DB 410 -----AFNPEPVLPPVSARPEQVEKVLKTRVHDATSKLSQGEID 449
QY 619 LMFIIISKRQLNAYGFVKHYCDHTIGVANQHITSETVTKALASLRHEKSKRIFYQIALK 678
DB 450 LLIVLPDNGSLYGLKRICETELGIVSQCLTKHVFKV-----SKQTMANVALK 500
QY 679 INAKLGGINQELDWSAEIAEISPEEKERRKTMPL-----TMVYGVDTVHTPTSYSGIDSYA 733
DB 501 INVKGGRNTVL-----VDALS-----RRILVSDRPTIIIFGADVTHHPGDSSESTA 549
QY 734 AVVASIN-PGGTIYRNMIVTQ-----EECRPGERAVAHGRERTDILEAKFVKLL 781
DB 550 AVVASQDWPEITYAGLVCAQAHQELIQDLFKWKDPQKGVVTG-----GMIKELL 601
QY 782 REFAENNDRAPAHIVYVRDGVSDSEMLRVSHDELRLSKSEVKQFMSEKDEGDEPPEKYTF 841
DB 602 IAFRRSTGHK-PLRIIFRYRDGVSEGFYQVLLLYELDAIR---KACASLBAGY--QPVPVF 655
QY 842 IVIQRHNTELLRRMEKPKPVNKNLTPAETDVAVAQVQWEEDMKESKETGINVSSGT 901
DB 656 VVQKRHHTRLFLAQNNHNRHSVDR-----SGNILP--GT 687
QY 902 TVDKLIVSKYKPDFFLASHHGVLTGRSPGHYTMVYDDKMGSDQEVYKMTYGLAFLSARCR 961

```



Db 688 VDSKICHPTEFDYLCSHAGIQTSPAHYHVLWDENNFTADGLQSLTNLCYTARCT 747  
QY 962 KPISLPVVPVYAHLSCEKAKELYRTYKEHYIGDYAQPRT 1000  
Db 748 RSVSIVPPAYYAHAAFRAR-----FVMEPET 774

RESULT 9  
US-10-449-902-56159  
; Sequence 56159, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 56159  
; LENGTH: 1011  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-56159

Query Match 9.3%; Score 500; DB 6; Length 1011;  
Best Local Similarity 22.0%; Pred. No. 2.3e-384;  
Matches 224; Conservative 172; Mismatches 236; Indels 236; Gaps 41;

QY 49 FSSKIYDREYVEYVMTKEVLRKPKGPPKTEPIPDRAKLFQHLRHEKQTDFIL 108  
Db 163 FFAHLPNKLHLDVDSITPEVTSR-----INVRAVI-----KELVNLK 201  
QY 109 EDVY-----FDEKDTVYSCRLNTVTSKMLVSEKVKVKKSEKKDKLEKKILYTMILT 162  
Db 202 ASVLGRPLPAYDGRKSLYTAGLPDPTSQEFQIT---LLDDDDGGSGRRQRTFRVVIKFA 258  
QY 163 YRKXPH---INFRENPEDKEANRSYKFLKNVMTQ---KVRYPAPFVNEEIKVQFAKNFVY 217  
Db 259 ARADLRHLRLEFLAGRAEAPQEA---LQVLDIVLRPLPSARYAPFGRSFPFSAVIGR--- 311  
QY 218 DNNISLRVPESFHDPRNFQOSLEVAPIRIEAWFGIYIGIKELFDGEPVLPFAIVDKLFYNA 277  
Db 312 -----RQP-----LGEGLSEWRGFGYSIRPTQMGSL-LNIDMSATAFIE- 349  
QY 278 PKMSLLDYLLLIIVDPQSCNDVVRKDKLTKLMAKMTIRQAAAPRIQLLENLKLCAEVW 337  
Db 350 -PUPVLDIV-----AQLNSDIH-----SRPLSDAERVKIKKALRGVKEVTH-- 391  
QY 338 DNEMSLRHLRFLDLCEBNSIYKYVTGKSDGRNAKYDTTLFKIYBENKXF-IEFPH 396  
Db 392 RGNMRR--KYRISGLTIQPRELTFPV-----DEGGTVK-----SVVQYFQETYGAIQHTY 441  
QY 397 LPLVTKYSAKEVAVPMHELVHEKQRYKNRIDLVMQDKFLKRAKRPKHDKYKENTLKM 456  
Db 442 LPCLTVQ---RLNLYPMVEVKIVE-GQRYSKRLNQIRALLBETQCHPRDRERDIKMW 497  
QY 457 KELDFSSBELFVERGLCSKLQMIQCPGVKLPML-VNSVNEQIKMTPVI---RGFQE 512  
Db 498 KNNAYQDDP---YAKEFGIKISDRLASVEARILPAFLKYNETGREKDCPLRVQNMNMK 555  
QY 513 KQLVVVPKELCCAVFVWNETAGNCPLEBNDVVVKFVTELIGGCKPFGI-----RIG 563  
Db 556 KVVNGGKVRSMVCVFNARN-----VQESVVRGFCHEALMCAQSGMDFAPFPIPLPLN 608

QY 564 ANENRGAQSMYDATKNEYAPYKNCINTLTGIGRPEIAATEAKNMPERLPDKEQKVLFI 623  
Db 609 AHPDQ-----VERALKARY-----HDMNVLG--PORRELDLIGI 642  
QY 624 ISKRQLNAVGFVHYCHDTIGVANQHITSETVTKALASLRHKGSKRIFVQIALKINAKL 683  
Db 643 LPDNGSLYGLDKRVCIEDLGIIVSCCCTKQVFM-----NKQILANLAKINVKV 693  
QY 684 GGINQELDWSEIARIISPEEKERRKTMPL-----TMVVGIDVTHPTSYSGIDSYIAAVVAS 738  
Db 694 GGRNTVL-----VDAYS-----RRIPLVTRPTIIFGADVTHPHPCGEDSSPSIAAVVAS 742  
QY 739 IN-PGCTIYRNMIQTQECRPERAVHGRE-----RTDILAEKFKVLREP 784  
Db 743 QDMPEVTKYAGLSAQ-----AHRQELIEDLYKIWDPOQRTVSGGMIRELLISF 792  
QY 785 AENNDRAPAHIVVYRDGVSDEMLRVSHDELSLSEVKQFMFSERDGEPEPKYTFIVI 844  
Db 793 KRSTGEK-PQRILIFRDGVSQGYQLLYELNAIRKACASLET-----NYQKVTIIV 846  
QY 845 QKQHNTLLRRMEKDKPVVKNKDLTPAETDVAAVAAVKQWEEDMKESKETGINVPSSGTTVD 904  
Db 847 QKQHNTLLFAHNNDQNSVDR-----SGNILP--GTVVD 878  
QY 905 KLIVSKYKTDFFLASHHGVLTSPRGHYTWYDDKMSQDEVYMTVGLAFLSARCKRPI 964  
Db 879 SKIRHPTFDFYLCSHAGINGTSRPAHYHVLWDENNFTADALQILTNLCYTVARCTRSV 938  
QY 965 SLFPVVPVYAHLSCEKAKELYRTYKEHYIGDYAQ-----PRTRHEMEHFLQTNVKYPG 1016  
Db 939 SIVPPAYYAHAAFR-----RFYMEPDTSDSSVSGVSGVGRGLSGSSTRTAPG 990

RESULT 10  
US-10-449-902-37356  
; Sequence 37356, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 37356  
; LENGTH: 876  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-37356

Query Match 8.1%; Score 431.5; DB 6; Length 876;  
Best Local Similarity 22.1%; Pred. No. 3.5e-19;  
Matches 223; Conservative 140; Mismatches 404; Indels 241; Gaps 37;

QY 26 RPTGKCDGKDYKFKVKKVLLVNVWPKFS-SKIYDREYVEYVMTKEV-LNRKPGKPPPKKTE 83  
Db 35 RPNLSEGMIGB-SCIVRTNCFVHLESDDQTIYEYDVCVTEVGINRAVIRELVK--- 90  
QY 84 IPIPRAKLFWOHLRHEKQTDPILEDYVDFEKOYTVSVCLNTVTSKMLVSEKVKKDS 143  
Db 91 -----QOKDSGLGRPLPAYDGRKRLYTSGLPFDHRLFLVLLDSIEDSP 134  
QY 144 EKKDEKDEKKILYTMILYRKFKHL-----NESRENPEKDEEANSRKFLKNVMTQ--KV 197  
Db 135 EBSRLRVRD---FVTLKFAAKISLWLRKFRGGKPNRESRA--ALRALDVVLKELPTA 189

```

QY 198 RVAPFVNEIEIKYQFAKFNVDNNSILRVPESPHDNRPQSLEVAPIEBAWFGVIGIKE 257
Db 190 RTQFAG-----SFSYN-LGECQOLCKVLESWGFQRIQA 225
QY 258 LFDGPEVLNFAIVDKLFYNAPKMSLLDYLLIVDPQSCNDDVRKOLTKLMAGKMTIRQA 317
Db 226 TQMGQLQ-LNIDVSSVFIK--PVPVDYV-----AQLLNEDILLD----- 262
QY 318 ARP-----RRQLLENILKCAEVDNEMSLRTERHLTFLDLCEENSLVYKVTGKSDR 370
Db 263 -RPLCSTEFKLKEALEGLKQI-----NGILENTYHVQDL 297
QY 371 GRNAKYDTTLFKIYEENKKEFEFHLPLVYKSGAKYAVPMHELVHEKQPKNRID 430
Db 298 VHQASFPVNF-----SIQYSLPCLKVNAHGETIFLPELVCKIAE-QCHQKQJLN 347
QY 431 LVMDQKFLKRAKRPKPHYKENTLKMELDFSEELNFERFGLCSLQMIKPCGKVLKE 490
Db 348 AXHMAALLQVAPPPNERDYNILQTVHONKY--QEDPHAKFEGIKIEEKLVSIKSRILPA 405
QY 491 PML-----VNSVNEQIKMTFVIRGFBQKQLNVVPEKELCCAVFV--VNETA-----GN 536
Db 406 PMLKFDHSDGETTEFLPQLGIMNMHKMIMGGRVKSACVNFCSVREYAAARNFCYDLGF 465
QY 537 PGLEENDV--VKFYTELI--GCKFRGIRIGANENRGAQSIWYDATKNEYAFYKNCITLN 591
Db 466 MCRESGWFSVAPVPLVTAKEGCVESALR-----TLH 498
QY 592 TGIGRFEIATAEAKMFERLPDKEQKVLMIIFIIISKRQLNAYGVFKYCHDTIGVANQHIT 651
Db 499 DDV-----MDILRPQGRKLDLILVILPNNNGSLYGDVKRICETDGLISQCCL 546
QY 652 SETVTYKALASLRHEKSGRIFFQIQAALKINAKLGGINQELDWSEIAEISPEEKERRKTML 711
Db 547 AKHVLKM-----NWKYLASVALKINAKMGGRNTVL-----VDALERMRLPHVRDTP-- 591
QY 712 TMYVGIDVTHPTSYSGIDYSIAAVVASIN-PGTYIRNMIVTQ-----EECRPGERAVAGR 767
Db 592 TIVGAHVTHPHGKANSSIAAVASQWPEVTKYAGISVQAQCHQESIQGLFKVQDDP 651
QY 768 ERTDILEAKFVKLLREFANNDRAPAHIVVYRDGVSDSEMLRSHDELRSLSKSEYKQPM 827
Db 652 ERGTTSGMIKEHLASFYBAT-KRKEGRIFIFRDGVSKQLPQALMHGELAIKMACASM- 709
QY 828 SERDEGDPKPTFVIVIQRHNTLRLRMEKDKPVVNKOLDTPAETDVAVAVKQHEEDMK 887
Db 710 -----GPDYNPLVTVYVLCRHTLRF-----ADYNNANTHDSANIR----- 747
QY 888 ESKETGINVPSSTGVDKLIVSKYDFDLASHHGVLGTSRPGHYTMVYDDKMGSDQEVY 947
Db 748 -----AGTVVDSNICQPNQDFLCLSHRSTQGTGRPRYTHVLWDENDFLAGSFQ 796
QY 948 KMTYGLAFLASRCRKPISLPVPHVAHLSCEKAKELYRTY-KEHYIGD 994
Db 797 ELTNVLCYTSATCTQSISVAVPHVYARLLSSRA-----RCYIKPRSIGD 840

```

```

RESULT 11
US-10-997-086-5
; Sequence 5, Application US/10997086
; Publication No. US20060135456A1
; GENERAL INFORMATION:
; APPLICANT: Hannon, Gregory J.
; APPLICANT: Paddison, Patrick J.
; APPLICANT: Siolas, Despina C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RNA
; TITLE OF INVENTION: INTERFERENCE
; FILE REFERENCE: CSHL-P05-010
; CURRENT APPLICATION NUMBER: US/10/997,086
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: US 10/350,798
; PRIOR FILING DATE: 2003-01-24

```

```

; PRIOR APPLICATION NUMBER: US 10/055,797
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/US01/08435
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 09/866,557
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/189,739
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: US 60/243,097
; PRIOR FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-997-086-5

```

```

Query Match      8.0%; Score 427; DB 6; Length 1145;
Best Local Similarity 21.1%; Pred. No. 9.8e-19;
Matches 210; Conservative 152; Mismatches 381; Indels 250; Gaps 37;

```

```

QY 59 YEYEVKMTYKVLNRKPGKPPKKTIPIDPRAKLFWHLRHEKKQTDFILEDYVDEKOT 118
Db 364 YHYDVKI-----MPEREKKEFYQAFEQFRVDQLGGAVLAYDGKAS 403
QY 119 VYSVCR--LNTVTSKMLYSE---KVVKKDSEKKDEKD--LEKKILYTMILTYRKKPHLNF 171
Db 404 CYSVDKLPUNSONPEVTVDNRGRTLYTIEIKETGDSIDLUKSLTYN----- 452
QY 172 SRENPEKDEEANRSYKFLKNVMTQKRVAPFVNBEIKVQFAKFNVDNNSILRVPESF-- 229
Db 453 ---NDRIFDKPMRAMQCVENVLASPC-----HNKAIRVGRSPFK 488
QY 230 -HDPNRFQSELEVAPIEAWFGIYIGIKELPDGPEVLNFAIVDKLFYNAPKMSLLDYLL 288
Db 489 MSDPNNRH---ELDDGYEALVGLYQAF--MLGDRPFLNVDI SHKSF--PISMPMIEYL-- 539
QY 289 IVDPOSCNDDVRKDLTKLMAGKMTIRQAARPIRQLLENLKLKCAEVDNEMSLRTERH 348
Db 540 -----ERFSLUKAKI--NNTNLDSYRRFLEPFLRGINV-----VTPPQS 577
QY 349 LTFDLCLBENSIVYKVTGSKDRNAKYD-----TTLFKIYEENKKEFEFHLPLVYKVK 403
Db 578 -----FQSAPRVYRVNGLSRAPASSETFEHDKGKVTIASYFHSRNYPLKFPQLHCLNVG 631
QY 404 SKAGEYAVPMHELVHEKPPQRYKNRIDLVMQ--DKFLKRAKRPKPHYKENTLKMELKELDFS 462
Db 632 SSIKSILLPIELCSIEE--GQALNRKDGATQVANNMIKYAATSTNVRKRIKIMNLLQYFOHN 689
QY 463 SEELNPFVERFGICSKLQMECPGKVLKBPMLVNSVNEQIKMTPIRVGFQE---KOLNVV 518
Db 690 LDPT--ISRFGIRIANDFIVVSTRVLSPPQV---EYHSKRFTWVKNVSGWRMDGMKLEBK 744
QY 519 PEKELCCAVFV-----VNETA---GNPCLSEENDVVKFYTELIGGCKFRGIRIGANEN 567
Db 745 PKAHKCAVLVCDPRSGRKNMYTQLNDPGLNLIISQCKAVNI--SLDSDVTVYRPF---TDDE 799
QY 568 RGAQSIMYDATKNEYAFYKNCITLNTGIRGFEIATAEAKMFERLPDKEQKVLMIIFIIISKR 627
Db 800 RSLDTIFADLKRSQHD-----LAIVIPQF 824
QY 628 QLNAYGVFKYCHDHTIGVANQHITSETVTYKALASLRHEKSGRIFFQIQAALKINAKLGGIN 687
Db 825 RI-SYDTTIKQKAELOHGILTOCIKQFTVER-----KCNNQTTGNILLKLNKUNGIN 875
QY 688 OELDWSEIAEISPEEKERRKKTMLTMYVGIDVTHPTSYSGIDYSIAAVVASINPGGTIYR 747
Db 876 HKI-----KDDPRLPMKNTMYIGADVTHPSDQORETPSVVGVAAASHDPYGCASYN 925
QY 748 NMIVTQECRPGERAVAGRERTDILEAKF---VKLLREFANNDRAPAHIVVYRDGV 804
Db 926 MQYRLQ-----RGALBIEIDWFSTLEHLRVYKEYR-NAYPDHIIYRDGV 971

```



; LENGTH: 458		
; TYPE: PRT		
; ORGANISM: Glycine max		
US-10-953-349-18474		
Query Match		7.5%; Score 403.5; DB 6; Length 458;
Best Local Similarity		28.1%; Pred. No. 7.3e-18;
Matches 121; Conservative 59; Mismatches 129; Indels 121; Gaps 14;		
QY	612	PKEQKV-----LMFIIISKRQLNAYGVKHYCDHTIGVANQHITSE 653
Db	60	PQVEKVLKTRYHDAKNKLQGRELDLLIVLPNNNGSLYGLDKRICETDLGLVSOCLTK 119
QY	654	TYTKALASLRHEKSGKRIFYQTALKINAKLGGINQELDMSEIAETISPEEKERRKTMPL-- 711
Db	120	HVFKM-----SKQYLANVALKINVKVGGRTVL-----VDALS-----RRIPLVS 159
QY	712	---TWVVGIDVTHPTSYSGIDYSIAAVASIN-PGGTIYRNMIVTQECPGERAVAHGR 767
Db	160	DRPTIIFGADVTHPHPGEDSSPSIAAVASQDYPEITKYAGLVCAQ-----VHRQ 209
QY	768	E-----RTDILEAKFVKLLREFAEENNDRAPAHIVVYRDGVSDESEMLRVSHD 814
Db	210	ELIQDLFKQWQDPVVRGTVTGGMIKELLISFRRATGOK-PQRIIFYRDGVSEGFYQVLLF 268
QY	815	ELRSLSKSEVKQPMSERDGEDPEPKY-----TFVIOKRHNTRLRRMEKDKPVVKNKDLTPA 870
Db	269	ELDAIRKACASL-----EPNYQPPVTFVVQKRHHTRLFASNHHDKSSVDK----- 314
QY	871	ETDVAVAVKQWEEDMKESKETGI VNPSSGTTVDKLI VSKYKFPDFFLASHHGVLTGTSRPG 930
Db	315	-----SGNILP--GTVVDSKI CHPTFEFDFVLC SHAGIQGTSRPA 351
QY	931	HYTMVDDKGMQDEVYKMTYGLAFLSARCRKPI SLPPVHYAHLSCSEKAKELYRTRYKEH 990
Db	352	HYHVLWDENNFTADALQTLTNLNCVYARCTRSVSIVPPAYYAHLAAPFRAR----- 402
QY	991	YIGDYAQPET 1000
Db	403	----FYMEPET 409

RESULT 15  
US-10-953-349-18473  
; Sequence 18473, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 18473  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-953-349-18473

Query Match  
Best Local Similarity 28.1%; Pred. No. 8.3e-18; Length 501;  
Matches 121; Conservative 59; Mismatches 129; Indels 121; Gaps 14;

QY	612	PKEQKV-----LMFIIISKRQLNAYGVKHYCDHTIGVANQHITSE 653
Db	103	PQVEKVLKTRYHDAKNKLQGRELDLLIVLPNNNGSLYGLDKRICETDLGLVSOCLTK 162
QY	654	TYTKALASLRHEKSGKRIFYQTALKINAKLGGINQELDMSEIAETISPEEKERRKTMPL-- 711
Db	163	HVFKM-----SKQYLANVALKINVKVGGRTVL-----VDALS-----RRIPLVS 202

Search completed: July 5, 2006, 13:50:09  
Job time : 29 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 14:11:38 ; Search time 23 Seconds  
(without alignments)  
1189.943 Million cell updates/sec

Title: US-10-645-746-3

Perfect score: 1020  
Sequence: 1 MSSNPFLEKGFYRHSIDPE.....RHEMHLQTNVYKPGMSFA 1020

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112942 seqs, 26832045 residues

Word size : 1

Total number of hits satisfying chosen parameters: 112925

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA New:

- 1: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 2: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 3: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 4: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 5: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 6: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 7: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*
- 8: /EMC Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	1.1	763	6	US-10-449-902-43163
2	11	1.1	861	7	US-11-192-437-4
3	9	0.9	94	7	US-11-192-437-13
4	9	0.9	94	7	US-11-192-437-14
5	9	0.9	94	7	US-11-192-437-15
6	9	0.9	471	6	US-10-953-349-19400
7	9	0.9	488	6	US-10-953-349-19399
8	9	0.9	527	6	US-10-953-349-19398
9	8	0.8	108	7	US-11-192-437-16
10	8	0.8	108	7	US-11-192-437-17
11	8	0.8	108	7	US-11-192-437-18
12	8	0.8	108	7	US-11-192-437-19
13	8	0.8	235	6	US-10-539-402-24
14	8	0.8	457	6	US-10-953-349-18475
15	8	0.8	458	6	US-10-953-349-18474
16	8	0.8	501	6	US-10-953-349-18473
17	8	0.8	514	6	US-10-953-349-2683
18	8	0.8	525	6	US-10-953-349-2682
19	8	0.8	547	6	US-10-953-349-2681
20	8	0.8	822	6	US-10-953-349-4066
21	8	0.8	870	6	US-10-953-349-4065
22	8	0.8	876	6	US-10-449-902-37356
23	8	0.8	1011	6	US-10-449-902-56159
24	8	0.8	1050	6	US-10-953-349-4064
25	8	0.8	1052	6	US-10-449-902-53188

26	7	0.7	103	6	US-10-449-902-43508	Sequence 43508, A
27	7	0.7	194	6	US-10-449-902-55668	Sequence 55668, A
28	7	0.7	216	6	US-10-449-902-43156	Sequence 43156, A
29	7	0.7	250	6	US-10-953-349-341	Sequence 341, Appl
30	7	0.7	271	7	US-11-330-363-15	Sequence 15, Appl
31	7	0.7	391	6	US-10-449-902-50154	Sequence 50154, A
32	7	0.7	436	6	US-10-449-902-40680	Sequence 40680, A
33	7	0.7	446	6	US-10-449-902-36916	Sequence 36916, A
34	7	0.7	465	6	US-10-449-902-45325	Sequence 45325, A
35	7	0.7	467	6	US-10-449-902-38097	Sequence 38097, A
36	7	0.7	477	6	US-10-449-902-38723	Sequence 38723, A
37	7	0.7	485	6	US-10-953-349-25267	Sequence 25267, A
38	7	0.7	489	6	US-10-449-902-54513	Sequence 54513, A
39	7	0.7	517	6	US-10-449-902-38630	Sequence 38630, A
40	7	0.7	518	6	US-10-449-902-37736	Sequence 37736, A
41	7	0.7	529	6	US-10-449-902-50447	Sequence 50447, A
42	7	0.7	551	7	US-11-236-238-29	Sequence 29, Appl
43	7	0.7	630	6	US-10-449-902-47718	Sequence 47718, A
44	7	0.7	632	6	US-10-449-902-48865	Sequence 48865, A
45	7	0.7	658	6	US-10-449-902-55772	Sequence 55772, A

#### ALIGNMENTS

##### RESULT 1

US-10-449-902-43163  
; Sequence 43163, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; PRIOR FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 43163  
; LENGTH: 763  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-43163

Query Match 1.1%; Score 11; DB 6; Length 763;  
Best Local Similarity 100.0%; Pred. No. 0.015;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687.

Db 612 LKINAKLGGIN 622

##### RESULT 2

US-11-192-437-4  
; Sequence 4, Application US/11192437  
; Publication No. US20060141600A1  
; GENERAL INFORMATION:  
; APPLICANT: Joshua-Tor, Leemor  
; APPLICANT: Song Ji-Joon  
; APPLICANT: Hannon, Gregory J.  
; APPLICANT: Liu, Jidong  
; APPLICANT: Carmell, Michelle A.  
; APPLICANT: Rivas, Fabiola  
; APPLICANT: Mareden, Carolyn  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO ARGONAUTE PROTEINS

```
; FILE REFERENCE: CSHL-P01-014
; CURRENT APPLICATION NUMBER: US/11/192,437
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US 60/592,297
; PRIOR FILING DATE: 2004-07-28
; PRIOR APPLICATION NUMBER: US 60/592,269
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-192-437-4
```

```
Query Match 1.1%; Score 11; DB 7; Length 861;
Best Local Similarity 100.0%; Pred. No. 0.016; 0; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0;
```

```
QY 677 LKINAKLGGIN 687
Db 557 LKINAKLGGIN 567
|||||
```

```
RESULT 3
US-11-192-437-13
```

```
; Sequence 13, Application US/11/192437
; Publication No. US20060141600A1
; GENERAL INFORMATION:
; APPLICANT: Joshua-Tor, Leemor
; APPLICANT: Song, Ji-Joon
; APPLICANT: Hannon, Gregory J.
; APPLICANT: Liu, Jidong
; APPLICANT: Carmell, Michelle A.
; APPLICANT: Rivas, Fabiola
; APPLICANT: Marsden, Carolyn
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO
; TITLE OF INVENTION: ARGONAUTE PROTEINS
; FILE REFERENCE: CSHL-P01-014
; CURRENT APPLICATION NUMBER: US/11/192,437
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US 60/592,297
; PRIOR FILING DATE: 2004-07-28
; PRIOR APPLICATION NUMBER: US 60/592,269
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Caenorhabditis briggsae
US-11-192-437-13
```

```
Query Match 0.9%; Score 9; DB 7; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.22; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
```

```
QY 796 IVVYRDGVS 804
Db 70 IVVYRDGVS 78
|||||
```

```
RESULT 4
```

```
US-11-192-437-14
; Sequence 14, Application US/11/192437
; Publication No. US20060141600A1
; GENERAL INFORMATION:
; APPLICANT: Joshua-Tor, Leemor
; APPLICANT: Song, Ji-Joon
; APPLICANT: Hannon, Gregory J.
; APPLICANT: Liu, Jidong
; APPLICANT: Carmell, Michelle A.
; APPLICANT: Rivas, Fabiola
```

```
; APPLICANT: Marsden, Carolyn
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO
; TITLE OF INVENTION: ARGONAUTE PROTEINS
; FILE REFERENCE: CSHL-P01-014
; CURRENT APPLICATION NUMBER: US/11/192,437
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US 60/592,297
; PRIOR FILING DATE: 2004-07-28
; PRIOR APPLICATION NUMBER: US 60/592,269
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-192-437-14
```

```
Query Match 0.9%; Score 9; DB 7; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.22; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
```

```
QY 796 IVVYRDGVS 804
Db 70 IVVYRDGVS 78
|||||
```

```
RESULT 5
```

```
US-11-192-437-15
; Sequence 15, Application US/11/192437
; Publication No. US20060141600A1
; GENERAL INFORMATION:
; APPLICANT: Joshua-Tor, Leemor
; APPLICANT: Song, Ji-Joon
; APPLICANT: Hannon, Gregory J.
; APPLICANT: Liu, Jidong
; APPLICANT: Carmell, Michelle A.
; APPLICANT: Rivas, Fabiola
; APPLICANT: Marsden, Carolyn
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO
; TITLE OF INVENTION: ARGONAUTE PROTEINS
; FILE REFERENCE: CSHL-P01-014
; CURRENT APPLICATION NUMBER: US/11/192,437
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US 60/592,297
; PRIOR FILING DATE: 2004-07-28
; PRIOR APPLICATION NUMBER: US 60/592,269
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-192-437-15
```

```
Query Match 0.9%; Score 9; DB 7; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.22; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
```

```
QY 796 IVVYRDGVS 804
Db 70 IVVYRDGVS 78
|||||
```

```
RESULT 6
```

```
US-10-953-349-19400
; Sequence 19400, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
```

```
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 19400
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-19400

Query Match          0.9%; Score 9; DB 6; Length 471;
Best Local Similarity 100.0%; Pred. No. 0.98;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGG 685
DB 172 LKINAKLGG 180

RESULT 7
US-10-953-349-19399
; Sequence 19399, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 19399
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-19399

Query Match          0.9%; Score 9; DB 6; Length 488;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGG 685
DB 189 LKINAKLGG 197

RESULT 8
US-10-953-349-19398
; Sequence 19398, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 19398
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-19398

Query Match          0.9%; Score 9; DB 6; Length 527;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGG 685
DB 189 LKINAKLGG 197
```

```
DB 228 LKINAKLGG 236

RESULT 9
US-11-192-437-16
; Sequence 16, Application US/11192437
; Publication No. US20060141600A1
; GENERAL INFORMATION:
; APPLICANT: Joshua-Tor, Leemor
; APPLICANT: Song, Ji-Joon
; APPLICANT: Hannon, Gregory J.
; APPLICANT: Liu, Jidong
; APPLICANT: Carmell, Michelle A.
; APPLICANT: Rivas, Fabiola
; APPLICANT: Marsden, Carolyn
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO
; FILE REFERENCE: CSHL-P01-014
; CURRENT APPLICATION NUMBER: US/11/192,437
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US 60/592,297
; PRIOR FILING DATE: 2004-07-28
; PRIOR APPLICATION NUMBER: US 60/592,269
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Oryza sativa
US-11-192-437-16

Query Match          0.8%; Score 8; DB 7; Length 108;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAUVAS 738
DB 16 SIAAUVAS 23

RESULT 10
US-11-192-437-17
; Sequence 17, Application US/11192437
; Publication No. US20060141600A1
; GENERAL INFORMATION:
; APPLICANT: Joshua-Tor, Leemor
; APPLICANT: Song, Ji-Joon
; APPLICANT: Hannon, Gregory J.
; APPLICANT: Liu, Jidong
; APPLICANT: Carmell, Michelle A.
; APPLICANT: Rivas, Fabiola
; APPLICANT: Marsden, Carolyn
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO
; FILE REFERENCE: CSHL-P01-014
; CURRENT APPLICATION NUMBER: US/11/192,437
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US 60/592,297
; PRIOR FILING DATE: 2004-07-28
; PRIOR APPLICATION NUMBER: US 60/592,269
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Oryza sativa
US-11-192-437-17

Query Match          0.8%; Score 8; DB 7; Length 108;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAUVAS 738
DB 16 SIAAUVAS 23
```



QY 731 SIAAVVAS 738  
| | | | | | | |  
DB 16 SIAAVVAS 23

RESULT 11  
US-11-192-437-18  
; Sequence 18, Application US/11192437  
; Publication No. US20060141600A1  
; GENERAL INFORMATION:  
; APPLICANT: Joshua-Tor, Leemor  
; APPLICANT: Hannon, Gregory J.  
; APPLICANT: Li, Jidong  
; APPLICANT: Carmell, Michelle A.  
; APPLICANT: Rivas, Fabiola  
; APPLICANT: Marsden, Carolyn  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO  
; TITLE OF INVENTION: ARGONAUTE PROTEINS  
; FILE REFERENCE: CSHL-P01-014  
; CURRENT APPLICATION NUMBER: US/11/192,437  
; CURRENT FILING DATE: 2005-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,297  
; PRIOR FILING DATE: 2004-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,269  
; PRIOR FILING DATE: 2004-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Arabidopsis  
US-11-192-437-18

Query Match 0.8%; Score 8; DB 7; Length 108;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAVVAS 738  
| | | | | | | |  
DB 16 SIAAVVAS 23

RESULT 12  
US-11-192-437-19  
; Sequence 19, Application US/11192437  
; Publication No. US20060141600A1  
; GENERAL INFORMATION:  
; APPLICANT: Joshua-Tor, Leemor  
; APPLICANT: Song, Ji-Joon  
; APPLICANT: Hannon, Gregory J.  
; APPLICANT: Li, Jidong  
; APPLICANT: Carmell, Michelle A.  
; APPLICANT: Rivas, Fabiola  
; APPLICANT: Marsden, Carolyn  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATED TO  
; TITLE OF INVENTION: ARGONAUTE PROTEINS  
; FILE REFERENCE: CSHL-P01-014  
; CURRENT APPLICATION NUMBER: US/11/192,437  
; CURRENT FILING DATE: 2005-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,297  
; PRIOR FILING DATE: 2004-07-28  
; PRIOR APPLICATION NUMBER: US 60/592,269  
; PRIOR FILING DATE: 2004-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Arabidopsis zwille  
US-11-192-437-19

Query Match 0.8%; Score 8; DB 7; Length 108;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAVVAS 738  
| | | | | | | |  
DB 16 SIAAVVAS 23

RESULT 13  
US-10-539-402-24  
; Sequence 24, Application US/10539402  
; Publication No. US20060115477A1  
; GENERAL INFORMATION:  
; APPLICANT: Xerion Pharmaceuticals AG  
; APPLICANT: Tufts University  
; TITLE OF INVENTION: Neurofilin-1 Inhibitor  
; FILE REFERENCE: XE12EPC  
; CURRENT APPLICATION NUMBER: US/10/539,402  
; CURRENT FILING DATE: 2005-06-17  
; PRIOR APPLICATION NUMBER: US 60/435,893  
; PRIOR FILING DATE: 2002-12-20  
; PRIOR APPLICATION NUMBER: EP 03000615  
; PRIOR FILING DATE: 2003-01-15  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: human  
US-10-539-402-24

Query Match 0.8%; Score 8; DB 6; Length 235;  
Best Local Similarity 100.0%; Pred. No. 5.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 ELSRLKSE 822  
| | | | | | | |  
DB 65 ELSRLKSE 72

RESULT 14  
US-10-953-349-18475  
; Sequence 18475, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 18475  
; LENGTH: 457  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-953-349-18475

Query Match 0.8%; Score 8; DB 6; Length 457;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAVVAS 738  
| | | | | | | |  
DB 181 SIAAVVAS 188

RESULT 15  
US-10-953-349-18474  
; Sequence 18474, Application US/10953349  
; Publication No. US20060107345A1

```
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579FUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 18474
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-18474

Query Match      0.8%; Score 8; DB 6; Length 458;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      731 SIAAVVAS 738
Db      182 SIAAVVAS 189
```

Search completed: July 5, 2006, 14:15:09  
Job time : 24 secs

THIS PAGE BLANK (USPTO)



```
Db 241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLLIVDPQSCNDDYR 300
QY 301 KDLKTKLMAGKMTIRQAARPRIQOLLENLKLKCAEVDNEMSRLLTERHLTFLDLCEENSL 360
Db 301 KDLKTKLMAGKMTIRQAARPRIQOLLENLKLKCAEVDNEMSRLLTERHLTFLDLCEENSL 360
QY 361 VYKVTGKSDRGNAKYDTTLFKIYEENKKEFEFFHLPLAVKVGSGAKAYAVPMHELVHE 420
Db 361 VYKVTGKSDRGNAKYDTTLFKIYEENKKEFEFFHLPLAVKVGSGAKAYAVPMHELVHE 420
QY 421 KPORYKNRIDLVMQDKFLKRAATRKPHDYKENTLKMKELDFFSSEELNLFVERFGLCSKLOM 480
Db 421 KPORYKNRIDLVMQDKFLKRAATRKPHDYKENTLKMKELDFFSSEELNLFVERFGLCSKLOM 480
QY 481 IECPGKVLKEPMLVNSVNEQIKMTPIVIRGFBQKQLNVVPEKELCCAFVFNNTAGNPCLE 540
Db 481 IECPGKVLKEPMLVNSVNEQIKMTPIVIRGFBQKQLNVVPEKELCCAFVFNNTAGNPCLE 540
QY 541 ENDVVKFYTELIGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600
Db 541 ENDVVKFYTELIGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600
QY 601 ATEAKMFERLPDKEQKVLMPFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTTKALA 660
Db 601 ATEAKMFERLPDKEQKVLMPFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTTKALA 660
QY 661 SLRHEKSGKRIFYQIALKINAKLGGINQELDSEIAEISPEBKERRKTMTPLTWYVGIDVT 720
Db 661 SLRHEKSGKRIFYQIALKINAKLGGINQELDSEIAEISPEBKERRKTMTPLTWYVGIDVT 720
QY 721 HPTSYSGIDYSTAAVVASINPGGTIYRNMIVTQEECRPGERAVAHGRETDILEAKFVKL 780
Db 721 HPTSYSGIDYSTAAVVASINPGGTIYRNMIVTQEECRPGERAVAHGRETDILEAKFVKL 780
QY 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKSEVKQPMSEDRDGEDPEPKYT 840
Db 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKSEVKQPMSEDRDGEDPEPKYT 840
QY 841 FVIOKRNHTRLLRRMEKDKPVVVKDLTPAETDVAVAVKQWEEDMKESKETGIVNPSGG 900
Db 841 FVIOKRNHTRLLRRMEKDKPVVVKDLTPAETDVAVAVKQWEEDMKESKETGIVNPSGG 900
QY 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTMVYDDKGMQSDDEVYKMTYGLAFLSARC 960
Db 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTMVYDDKGMQSDDEVYKMTYGLAFLSARC 960
```

## RESULT 2

```
US-10-645-735-3
; Sequence 3, Application US/10645735
; Publication No. US20050100913A1
; GENERAL INFORMATION:
; APPLICANT: Mello, Craig C.
; APPLICANT: Tabara, Hiroaki
; APPLICANT: Grishok, Alla
; APPLICANT: Fire, Andrew
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC
; TITLE OF INVENTION: INTERFERENCE
; FILE REFERENCE: UMG-052
; CURRENT APPLICATION NUMBER: US/10/645,735
; CURRENT FILING DATE: 2003-08-20
; PRIOR APPLICATION NUMBER: US/09/689,992A
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 3
; LENGTH: 1020
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-645-735-3
```

## Query Match

100.0%; Score 1020; DB 5; Length 1020;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNPPELEKGFYRHSLOPEMKWLARPCTGKCDGKFYEKKVLLLVNMFKSSKIYDREYVE 60

Db 1 MSSNPPELEKGFYRHSLOPEMKWLARPCTGKCDGKFYEKKVLLLVNMFKSSKIYDREYVE 60

QY 61 YEYKMTKEVLRNPKPKPPKTEIPIPORAKLFWOHLRHEKKQTPTILEDYVDFEKTQY 120

Db 61 YEYKMTKEVLRNPKPKPPKTEIPIPORAKLFWOHLRHEKKQTPTILEDYVDFEKTQY 120

QY 121 SVCRLLNTVTSKMLVSEKVKVDSEKKDKDEKLEKKILYTMILTYRKFFHLNFRSRENEPKDE 180

Db 121 SVCRLLNTVTSKMLVSEKVKVDSEKKDKDEKLEKKILYTMILTYRKFFHLNFRSRENEPKDE 180

QY 181 EARNRSYKFLKNVMTQKVRYAPFVNNEEIKVQFAKNFVYDNNNSILRVPESPHDPNRRPEQSLE 240

Db 181 EARNRSYKFLKNVMTQKVRYAPFVNNEEIKVQFAKNFVYDNNNSILRVPESPHDPNRRPEQSLE 240

QY 241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLLIVDPQSCNDDYR 300

Db 241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLLIVDPQSCNDDYR 300

QY 301 KDLKTKLMAGKMTIRQAARPRIQOLLENLKLKCAEVDNEMSRLLTERHLTFLDLCEENSL 360

Db 301 KDLKTKLMAGKMTIRQAARPRIQOLLENLKLKCAEVDNEMSRLLTERHLTFLDLCEENSL 360

QY 361 VYKVTGKSDRGNAKYDTTLFKIYEENKKEFEFFHLPLAVKVGSGAKAYAVPMHELVHE 420

Db 361 VYKVTGKSDRGNAKYDTTLFKIYEENKKEFEFFHLPLAVKVGSGAKAYAVPMHELVHE 420

QY 421 KPORYKNRIDLVMQDKFLKRAATRKPHDYKENTLKMKELDFFSSEELNLFVERFGLCSKLOM 480

Db 421 KPORYKNRIDLVMQDKFLKRAATRKPHDYKENTLKMKELDFFSSEELNLFVERFGLCSKLOM 480

QY 481 IECPCVKLKEPMLVNSVNEQIKMTPIVIRGFBQKQLNVVPEKELCCAFVFNNTAGNPCLE 540

Db 481 IECPCVKLKEPMLVNSVNEQIKMTPIVIRGFBQKQLNVVPEKELCCAFVFNNTAGNPCLE 540

QY 541 ENDVVKFYTELIGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600

Db 541 ENDVVKFYTELIGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600

QY 601 ATEAKMFERLPDKEQKVLMPFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTTKALA 660

Db 601 ATEAKMFERLPDKEQKVLMPFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTTKALA 660

QY 661 SLRHEKSGKRIFYQIALKINAKLGGINQELDSEIAEISPEBKERRKTMTPLTWYVGIDVT 720

Db 661 SLRHEKSGKRIFYQIALKINAKLGGINQELDSEIAEISPEBKERRKTMTPLTWYVGIDVT 720

QY 721 HPTSYSGIDYSTAAVVASINPGGTIYRNMIVTQEECRPGERAVAHGRETDILEAKFVKL 780

Db 721 HPTSYSGIDYSTAAVVASINPGGTIYRNMIVTQEECRPGERAVAHGRETDILEAKFVKL 780

QY 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKSEVKQPMSEDRDGEDPEPKYT 840

Db 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKSEVKQPMSEDRDGEDPEPKYT 840

QY 841 FVIOKRNHTRLLRRMEKDKPVVVKDLTPAETDVAVAVKQWEEDMKESKETGIVNPSGG 900

Db 841 FVIOKRNHTRLLRRMEKDKPVVVKDLTPAETDVAVAVKQWEEDMKESKETGIVNPSGG 900

QY 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTMVYDDKGMQSDDEVYKMTYGLAFLSARC 960

Db 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTMVYDDKGMQSDDEVYKMTYGLAFLSARC 960

Qy 961 RKPISLPVPVHYAHLSCAKELRYTYKEHYIGDYAOPRTRHEHEHFLQTNVVKPGMSFA 1020  
Db 961 RKPISLPVPVHYAHLSCAKELRYTYKEHYIGDYAOPRTRHEHEHFLQTNVVKPGMSFA 1020

RESULT 3  
US-11-144-985-3  
; Sequence 3, Application US/11144985  
; Publication No. US20060024798A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; FILE REFERENCE: UMY-052CN  
; CURRENT APPLICATION NUMBER: US/11/144,985  
; PRIOR FILING DATE: 2005-06-03  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1020  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans

Query Match 100.0%; Score 1020; DB 6; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSNFPLEKGFYRHSIDPEMKMLARTGKDGKFKYKCVLLLVNWFKFSKIYDREYVE 60  
Db 1 MSNFPLEKGFYRHSIDPEMKMLARTGKDGKFKYKCVLLLVNWFKFSKIYDREYVE 60

Qy 61 YEVMTKEVLRKPGKFPKKTETIPIDRAKLFQHLRHEKQDFLEDYVFEKDTVY 120  
Db 61 YEVMTKEVLRKPGKFPKKTETIPIDRAKLFQHLRHEKQDFLEDYVFEKDTVY 120

Qy 121 SVCRNLNTVTSKMLVSEKVVKKDSEKDEKLEKKILYTMILTYRKKFHLNFSRENPEKDE 180  
Db 121 SVCRNLNTVTSKMLVSEKVVKKDSEKDEKLEKKILYTMILTYRKKFHLNFSRENPEKDE 180

Qy 181 EARSYKFLQNVMTQKVYAPFVNEEIKVQFAPKVFYDNNNSILRVPSFHDNRFQESLE 240  
Db 181 EARSYKFLQNVMTQKVYAPFVNEEIKVQFAPKVFYDNNNSILRVPSFHDNRFQESLE 240

Qy 241 VAPRIEAWFGIYIGIKELFGEVLPNFAIVDKLFPYNAKMSLLDYLLIYVDPQSCNDVVR 300  
Db 241 VAPRIEAWFGIYIGIKELFGEVLPNFAIVDKLFPYNAKMSLLDYLLIYVDPQSCNDVVR 300

Qy 301 KDLKTKLMAGKMTIROAARPRIQOLLENKLCAEVWDNEMSLRTERHLTFDLCEENSL 360  
Db 301 KDLKTKLMAGKMTIROAARPRIQOLLENKLCAEVWDNEMSLRTERHLTFDLCEENSL 360

Qy 361 VYKVTGSDGRNAKKYDTTLFKIYENKKFIFPPHPLPVKVKSGAKKEYAVPMHEHLEVE 420  
Db 361 VYKVTGSDGRNAKKYDTTLFKIYENKKFIFPPHPLPVKVKSGAKKEYAVPMHEHLEVE 420

Qy 421 KPORYKNRIDLVMQDKELKATRKPHDYKENTLQMLKELDFSSSEELNPFVRFGLCSLQW 480  
Db 421 KPORYKNRIDLVMQDKELKATRKPHDYKENTLQMLKELDFSSSEELNPFVRFGLCSLQW 480

Qy 481 IBCPGKVLKEPMLVNSNEQIKMTFVIRGFOEQKQNVVPEKELCCAVFVNVNAGNCPLE 540  
Db 481 IBCPGKVLKEPMLVNSNEQIKMTFVIRGFOEQKQNVVPEKELCCAVFVNVNAGNCPLE 540

Qy 541 ENDVVVKPYTELIGCKPGRGIRIGANENRGAOSIMYDATKNBYAFYKNCITLNTGIRPRIA 600  
Db 541 ENDVVVKPYTELIGCKPGRGIRIGANENRGAOSIMYDATKNBYAFYKNCITLNTGIRPRIA 600

Qy 601 ATEAKNMFERLPDKEQKVLMPFIIISKQLNAYGVFKHYCDHTIGVANOHIITSETVTKALA 660  
Db 601 ATEAKNMFERLPDKEQKVLMPFIIISKQLNAYGVFKHYCDHTIGVANOHIITSETVTKALA 660

Qy 661 SLRHEKSGKRIFYQIALKINAKLGGINQELDWSIAEISPEEKERRKTMPLTMYVGIDVT 720  
Db 661 SLRHEKSGKRIFYQIALKINAKLGGINQELDWSIAEISPEEKERRKTMPLTMYVGIDVT 720

Qy 721 HPTSYSGIDYSIAAVASINPGGTIYRNMIYVQECRPERGAVAHGRERTDILEAKFVKL 780  
Db 721 HPTSYSGIDYSIAAVASINPGGTIYRNMIYVQECRPERGAVAHGRERTDILEAKFVKL 780

Qy 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLRVSHDELRLSKSEVKQFMSERGEDDEPEPKYT 840  
Db 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLRVSHDELRLSKSEVKQFMSERGEDDEPEPKYT 840

Qy 841 FVIOQKEHNTLLRMEKDKPVPVNNKOLTPAETDVAVAVKQWEEDMKESKETGI VNPSSG 900  
Db 841 FVIOQKEHNTLLRMEKDKPVPVNNKOLTPAETDVAVAVKQWEEDMKESKETGI VNPSSG 900

Qy 901 TTVDKLI VSKYKDFFLASHHGVLTGRPGHYTVYDDKQMSQDEVYKQTYGLAFLSARC 960  
Db 901 TTVDKLI VSKYKDFFLASHHGVLTGRPGHYTVYDDKQMSQDEVYKQTYGLAFLSARC 960

Qy 961 RKPTSLPVPVHYAHLSCAKELRYTYKEHYIGDYAOPRTRHEHEHFLQTNVVKPGMSFA 1020  
Db 961 RKPTSLPVPVHYAHLSCAKELRYTYKEHYIGDYAOPRTRHEHEHFLQTNVVKPGMSFA 1020

RESULT 4  
US-10-645-746-13  
; Sequence 13, Application US/10645746  
; Publication No. US20040265839A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; FILE REFERENCE: UMY-052DV1  
; CURRENT APPLICATION NUMBER: US/10/645,746  
; PRIOR FILING DATE: 2003-08-20  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 818  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans

Query Match 80.2%; Score 818; DB 5; Length 818;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 VNEEIKVQFAKNFVYDNNNSILRVPSFHDNRFQESLEAVAPRIEAWFGIYIGIKELFGE 262  
Db 1 VNEEIKVQFAKNFVYDNNNSILRVPSFHDNRFQESLEAVAPRIEAWFGIYIGIKELFGE 60

Qy 263 PVLNFAIVDKLFPYNAKMSLLDYLLIYVDPQSCNDVVRKDLKTKLMAGKMTIROAARPRI 322  
Db 61 PVLNFAIVDKLFPYNAKMSLLDYLLIYVDPQSCNDVVRKDLKTKLMAGKMTIROAARPRI 120

```
QY 323 ROLLENLKLKCAEVDNENSLRTERHLTFLDLCEENSLVYKVTGKSDRGNAKKYDTTLF 382
DB 121 ROLLENLKLKCAEVDNENSLRTERHLTFLDLCEENSLVYKVTGKSDRGNAKKYDTTLF 180
QY 383 KIYEENKKEIEPPHPLPLVKVKGAKAYAVPMEHLEVHEKPORYKQRIIDLVMQDKFLKRAT 442
DB 181 KIYEENKKEIEPPHPLPLVKVKGAKAYAVPMEHLEVHEKPORYKQRIIDLVMQDKFLKRAT 240
QY 443 RKPHDYKENTLKMELDFSSBELNFVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 502
DB 241 RKPHDYKENTLKMELDFSSBELNFVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 300
QY 503 MTPVIRGFOEKQLNVVPEKELCCAVFVNENAGNCLCEENDVVKFYTELIGCKPGRIRI 562
DB 301 MTPVIRGFOEKQLNVVPEKELCCAVFVNENAGNCLCEENDVVKFYTELIGCKPGRIRI 360
QY 563 GANENRGAQSIYDATKNEYAFYKNCITLNTGIRGFEIAATEAKNMPERLPDKEQKVLMEI 622
DB 361 GANENRGAQSIYDATKNEYAFYKNCITLNTGIRGFEIAATEAKNMPERLPDKEQKVLMEI 420
QY 623 IISKQQLNAYGFVKHYCDHTIGVANOHTITSETVTKALASLRHEKSGSKRIFYQIALKINAK 682
DB 421 IISKQQLNAYGFVKHYCDHTIGVANOHTITSETVTKALASLRHEKSGSKRIFYQIALKINAK 480
QY 683 LGGINQELDWSIAETISPEEKERRKTMPLTMVYGVIDVTHPTSYSGIDYSIAAVVASINPG 742
DB 481 LGGINQELDWSIAETISPEEKERRKTMPLTMVYGVIDVTHPTSYSGIDYSIAAVVASINPG 540
QY 743 GTIYRNMIVTQBECPGERAVAGHRETDILEAKFVKLLREFAENNDRAPAHIVVYRDG 802
DB 541 GTIYRNMIVTQBECPGERAVAGHRETDILEAKFVKLLREFAENNDRAPAHIVVYRDG 600
QY 803 VSDSEMLRVSHDELRSLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 862
DB 601 VSDSEMLRVSHDELRSLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 660
QY 863 VNKDLTPAETDVAAVAAVKQWEDMKESKETGIVNPSGGTTVDKLIYSKYKFFFLASHHG 922
DB 661 VNKDLTPAETDVAAVAAVKQWEDMKESKETGIVNPSGGTTVDKLIYSKYKFFFLASHHG 720
QY 923 VLGTSRPGHYTWMYDDKQMSQDEVYKMTYGLAFLSARCKRPISLPVPVHYAHLSCCKAKE 982
DB 721 VLGTSRPGHYTWMYDDKQMSQDEVYKMTYGLAFLSARCKRPISLPVPVHYAHLSCCKAKE 780
QY 983 LYRTYKEHYIGDYAOPRTRHEMEHFLQTNVKYPGMSFA 1020
DB 781 LYRTYKEHYIGDYAOPRTRHEMEHFLQTNVKYPGMSFA 818
```

## RESULT 5

US-10-645-735-13

; Sequence 13, Application US/10645735

; Publication No. US20050100913A1

; GENERAL INFORMATION:

; APPLICANT: Mello, Craig C.

; APPLICANT: Tabara, Hiroaki

; APPLICANT: Grishok, Alla

; APPLICANT: Fire, Andrew

; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC

; TITLE OF INVENTION: INTERFERENCE

; FILE REFERENCE: UMG-052

; CURRENT APPLICATION NUMBER: US/10/645,735

; CURRENT FILING DATE: 2003-08-20

; PRIOR APPLICATION NUMBER: US/09/689,992A

; PRIOR FILING DATE: 2000-10-13

; PRIOR APPLICATION NUMBER: US 60/193,218

; PRIOR FILING DATE: 2000-03-30

; PRIOR APPLICATION NUMBER: US 60/159,776

; PRIOR FILING DATE: 1999-10-15

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 818

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-10-645-735-13

Query Match 80.2%; Score 818; DB 5; Length 818;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 203 VNEEIKVQFAKNFVYDNNISILRVPSFHDPNRPFOSLEVAAPRIEAWFGIYIGIKELFDGE 262
DB 1 VNEEIKVQFAKNFVYDNNISILRVPSFHDPNRPFOSLEVAAPRIEAWFGIYIGIKELFDGE 60
QY 263 PVLNFAIVDKLFPYNAKMSLLDYLLIYVDPQSCNDVVRKDLATKLMAGKWTIRQAAAPRI 322
DB 61 PVLNFAIVDKLFPYNAKMSLLDYLLIYVDPQSCNDVVRKDLATKLMAGKWTIRQAAAPRI 120
QY 323 ROLLENLKLKCAEVDNENSLRTERHLTFLDLCEENSLVYKVTGKSDRGNAKKYDTTLF 382
DB 121 ROLLENLKLKCAEVDNENSLRTERHLTFLDLCEENSLVYKVTGKSDRGNAKKYDTTLF 180
QY 383 KIYEENKKEIEPPHPLPLVKVKGAKAYAVPMEHLEVHEKPORYKQRIIDLVMQDKFLKRAT 442
DB 181 KIYEENKKEIEPPHPLPLVKVKGAKAYAVPMEHLEVHEKPORYKQRIIDLVMQDKFLKRAT 240
QY 443 RKPHDYKENTLKMELDFSSBELNFVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 502
DB 241 RKPHDYKENTLKMELDFSSBELNFVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 300
QY 503 MTPVIRGFOEKQLNVVPEKELCCAVFVNENAGNCLCEENDVVKFYTELIGCKPGRIRI 562
DB 301 MTPVIRGFOEKQLNVVPEKELCCAVFVNENAGNCLCEENDVVKFYTELIGCKPGRIRI 360
QY 563 GANENRGAQSIYDATKNEYAFYKNCITLNTGIRGFEIAATEAKNMPERLPDKEQKVLMEI 622
DB 361 GANENRGAQSIYDATKNEYAFYKNCITLNTGIRGFEIAATEAKNMPERLPDKEQKVLMEI 420
QY 623 IISKQQLNAYGFVKHYCDHTIGVANOHTITSETVTKALASLRHEKSGSKRIFYQIALKINAK 682
DB 421 IISKQQLNAYGFVKHYCDHTIGVANOHTITSETVTKALASLRHEKSGSKRIFYQIALKINAK 480
QY 683 LGGINQELDWSIAETISPEEKERRKTMPLTMVYGVIDVTHPTSYSGIDYSIAAVVASINPG 742
DB 481 LGGINQELDWSIAETISPEEKERRKTMPLTMVYGVIDVTHPTSYSGIDYSIAAVVASINPG 540
QY 743 GTIYRNMIVTQBECPGERAVAGHRETDILEAKFVKLLREFAENNDRAPAHIVVYRDG 802
DB 541 GTIYRNMIVTQBECPGERAVAGHRETDILEAKFVKLLREFAENNDRAPAHIVVYRDG 600
QY 803 VSDSEMLRVSHDELRSLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 862
DB 601 VSDSEMLRVSHDELRSLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 660
QY 863 VNKDLTPAETDVAAVAAVKQWEDMKESKETGIVNPSGGTTVDKLIYSKYKFFFLASHHG 922
DB 661 VNKDLTPAETDVAAVAAVKQWEDMKESKETGIVNPSGGTTVDKLIYSKYKFFFLASHHG 720
QY 923 VLGTSRPGHYTWMYDDKQMSQDEVYKMTYGLAFLSARCKRPISLPVPVHYAHLSCCKAKE 982
DB 721 VLGTSRPGHYTWMYDDKQMSQDEVYKMTYGLAFLSARCKRPISLPVPVHYAHLSCCKAKE 780
QY 983 LYRTYKEHYIGDYAOPRTRHEMEHFLQTNVKYPGMSFA 1020
DB 781 LYRTYKEHYIGDYAOPRTRHEMEHFLQTNVKYPGMSFA 818
```

## RESULT 6

US-11-144-985-13

; Sequence 13, Application US/11144985

; Publication No. US20060024798A1

; GENERAL INFORMATION:

; APPLICANT: Mello, Craig C.

; APPLICANT: Tabara, Hiroaki



```
; APPLICANT: Grishok, Alla
; APPLICANT: Fire, Andrew
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC
; TITLE OF INVENTION: INTERFERENCE
; FILE REFERENCE: UMY-052CN
; CURRENT APPLICATION NUMBER: US/11/144,985
; PRIOR FILING DATE: 2005-06-03
; PRIOR APPLICATION NUMBER: US 09/689,992
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-144-985-13

Query Match      80.2%; Score 818; DB 6; Length 818;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 VNEBIKQVAKNFYDNNILRVDPESHDPNRFQSLVAPRIEAWFGIYIGIKELPDGE 262
DB 1 VNEBIKQVAKNFYDNNILRVDPESHDPNRFQSLVAPRIEAWFGIYIGIKELPDGE 60

QY 263 PVLNFAIVDKLFTYNAPKMSLDYLLLVDPQSCNDDYKDKTKMLMAGKMTIRQAAPRI 322
DB 61 PVLNFAIVDKLFTYNAPKMSLDYLLLVDPQSCNDDYKDKTKMLMAGKMTIRQAAPRI 120

QY 323 ROLLENLKLKCAEVDNEMSLRTERHLTFLDLCEENSLVYKVTGKSDRGNAKKYDTTLF 382
DB 121 ROLLENLKLKCAEVDNEMSLRTERHLTFLDLCEENSLVYKVTGKSDRGNAKKYDTTLF 180

QY 383 KIYEENKFTFPHLPLVYKSGAKYAVPMEHLEVEHKEPQRYKNRIDLVQDKFLKRAI 442
DB 181 KIYEENKFTFPHLPLVYKSGAKYAVPMEHLEVEHKEPQRYKNRIDLVQDKFLKRAI 240

QY 443 RKPHDYKENTLKMELDPSSBELFVERFGLCKLQMIKCPGKVLKEPMLVNSVNEQIK 502
DB 241 RKPHDYKENTLKMELDPSSBELFVERFGLCKLQMIKCPGKVLKEPMLVNSVNEQIK 300

QY 503 MTPVIRGFQSKQLNVVPEKELCCAVFVNNETAGNPLCEENDVVKFYTELIGGCKFRGIRI 562
DB 301 MTPVIRGFQSKQLNVVPEKELCCAVFVNNETAGNPLCEENDVVKFYTELIGGCKFRGIRI 360

QY 563 GANENRGAQSIMYDQKNEVAFYKNCYTLNTGIGRFEIAATEAKNMPERLPDKSQKVLMPFI 622
DB 361 GANENRGAQSIMYDQKNEVAFYKNCYTLNTGIGRFEIAATEAKNMPERLPDKSQKVLMPFI 420

QY 623 IISKROLNAGFVKHYCDHTIGVANOHITSETVTKALASIRHEKSGKRIFYQIALKINAK 682
DB 421 IISKROLNAGFVKHYCDHTIGVANOHITSETVTKALASIRHEKSGKRIFYQIALKINAK 480

QY 683 LGGINQELDWSEIAEISPEEKERRKTMPLTMYGIDVTHPTSYSGIDYSIAAUVASINPG 742
DB 481 LGGINQELDWSEIAEISPEEKERRKTMPLTMYGIDVTHPTSYSGIDYSIAAUVASINPG 540

QY 743 GTTYRNMIVTQESCRPGERAVAGRERTDILEAKFVKLLREFAENNDRNAPAHIVVYRDG 802
DB 541 GTTYRNMIVTQESCRPGERAVAGRERTDILEAKFVKLLREFAENNDRNAPAHIVVYRDG 600

QY 803 VSDSEMLRVSHDELRSIKSVKQPMSEKEDGEPKPYTFIVIOKRNTRILRMEWKQPV 862
DB 601 VSDSEMLRVSHDELRSIKSVKQPMSEKEDGEPKPYTFIVIOKRNTRILRMEWKQPV 660

QY 863 VNKDLPATETDVAVAQKQWEEDMKESKETGIWNPSSGTTVDKLIYSKYKFDFFLASHHG 922
DB 661 VNKDLPATETDVAVAQKQWEEDMKESKETGIWNPSSGTTVDKLIYSKYKFDFFLASHHG 720
```

```
QY 923 VLGTSRPGHYTMVYDDKMGMSQDEVYKMTYGLAFISARCRKPISLPVPVHYAHLSCKEKAKE 982
DB 721 VLGTSRPGHYTMVYDDKMGMSQDEVYKMTYGLAFISARCRKPISLPVPVHYAHLSCKEKAKE 780

QY 983 LYRTYKEHYIGDYAQPRTHEMEHFIQTNVYKPGMSFA 1020
DB 781 LYRTYKEHYIGDYAQPRTHEMEHFIQTNVYKPGMSFA 818

RESULT 7
US-10-437-963-131209
; Sequence 131209, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 131209
; LENGTH: 820
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_33297C.1.pep
US-10-437-963-131209

Query Match      1.1%; Score 11; DB 4; Length 820;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 574 LKINAKLGGIN 584

RESULT 8
US-10-437-963-131210
; Sequence 131210, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 131210
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(901)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
```

```
; OTHER INFORMATION: Clone ID: PAT_MRT4530_33298C.1.pep
US-10-437-963-131210

Query Match          1.1%; Score 11; DB 4; Length 901;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 602 LKINAKLGGIN 612

RESULT 9
US-10-408-765A-995
; Sequence 995, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 995
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-995

Query Match          1.1%; Score 11; DB 4; Length 924;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 620 LKINAKLGGIN 630

RESULT 10
US-10-408-765A-1792
; Sequence 1792, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1792
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1792

Query Match          1.0%; Score 10; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.7;

; OTHER INFORMATION: Clone ID: PAT_MRT4530_33298C.1.pep
US-10-437-963-131210

Query Match          1.1%; Score 11; DB 4; Length 901;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 602 LKINAKLGGIN 612

RESULT 9
US-10-408-765A-995
; Sequence 995, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 995
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-995

Query Match          1.1%; Score 11; DB 4; Length 924;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 620 LKINAKLGGIN 630

RESULT 10
US-10-408-765A-1792
; Sequence 1792, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1792
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1792

Query Match          1.0%; Score 10; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.7;

; OTHER INFORMATION: Clone ID: LIB83-005-F5_FLI.pep
US-10-425-114-37286

Query Match          1.0%; Score 10; DB 4; Length 678;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 STAAVVASIN 740
DB 440 STAAVVASIN 449

RESULT 11
US-09-801-574-16
; Sequence 16, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijing Jeremy
; APPLICANT: Page, David C.
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-801-574-16

Query Match          1.0%; Score 10; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVSD 805
DB 417 IVVYRDGVSD 426

RESULT 12
US-10-425-114-37286
; Sequence 37286, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 37286
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB83-005-F5_FLI.pep
US-10-425-114-37286

Query Match          1.0%; Score 10; DB 4; Length 678;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 STAAVVASIN 740
DB 440 STAAVVASIN 449
```

## RESULT 13

US-10-424-599-244834  
; Sequence 244834, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 244834  
; LENGTH: 117  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_63114C.1.pep  
US-10-424-599-244834

Query Match 0.9%; Score 9; DB 4; Length 117;  
Best Local Similarity 100.0%; Pred. No. 4.2; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 677 LKINAKLGG 685  
|||||  
Db 60 LKINAKLGG 68

## RESULT 14

US-10-425-115-278723  
; Sequence 278723, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 278723  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(303)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_185770C.1.pep  
US-10-425-115-278723

Query Match 0.9%; Score 9; DB 4; Length 303;  
Best Local Similarity 100.0%; Pred. No. 10; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 677 LKINAKLGG 685  
|||||  
Db 5 LKINAKLGG 13

## RESULT 15

US-10-174-363-50  
; Sequence 50, Application US/10174363  
; Publication No. US20030077623A1  
; GENERAL INFORMATION:

; APPLICANT: Butler, Karlene H.  
; APPLICANT: Cahoon, Rebecca E  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Rafaleki, Antoni J.  
; APPLICANT: Sakai, Hajime  
; TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptional  
; TITLE OF INVENTION: Gene Silencing  
; FILE REFERENCE: BB1454 US NA  
; CURRENT APPLICATION NUMBER: US/10/174,363  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: 60/298,973  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 50  
; LENGTH: 389  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-174-363-50  
  
Query Match 0.9%; Score 9; DB 4; Length 389;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 677 LKINAKLGG 685  
|||||  
Db 89 LKINAKLGG 97  
  
Search completed: July 5, 2006, 14:14:42  
Job time : 193 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 13:59:07 ; Search time 52 Seconds  
(without alignments)  
1716.947 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 1020  
Sequence: 1 MSSNPFLKGFYRHSLDPR.....RHEMFLOQNVKYPGMSFA 1020

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 650591 seqs, 87530628 residues

Word size : 1

Total number of hits satisfying chosen parameters: 649417

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC Celerra\_SID33/ptodata/2/iaa/5 COMB.pep:\*
- 2: /EMC Celerra\_SID33/ptodata/2/iaa/6 COMB.pep:\*
- 3: /EMC Celerra\_SID33/ptodata/2/iaa/7 COMB.pep:\*
- 4: /EMC Celerra\_SID33/ptodata/2/iaa/H COMB.pep:\*
- 5: /EMC Celerra\_SID33/ptodata/2/iaa/PCTUS COMB.pep:\*
- 6: /EMC Celerra\_SID33/ptodata/2/iaa/RE COMB.pep:\*
- 7: /EMC Celerra\_SID33/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	0.8	179	2	US-09-270-767-40274
2	8	0.8	179	2	US-09-270-767-55490
3	8	0.8	243	1	US-08-685-992-35
4	8	0.8	243	1	US-09-144-925-35
5	8	0.8	445	2	US-09-328-352-6669
6	8	0.8	452	2	US-10-332-795-7
7	8	0.8	969	2	US-09-533-029-106
8	7	0.7	15	1	US-08-368-251-19
9	7	0.7	15	5	PCT-US93-01112-19
10	7	0.7	23	1	US-08-086-428B-159
11	7	0.7	23	1	US-08-268-251-49
12	7	0.7	23	1	US-08-468-570-159
13	7	0.7	23	1	US-08-290-665A-263
14	7	0.7	23	2	US-08-466-601A-159
15	7	0.7	23	5	PCT-US93-01112-19
16	7	0.7	23	5	PCT-US95-10398-263
17	7	0.7	60	2	US-09-248-796A-25158
18	7	0.7	77	1	US-08-774-065-6
19	7	0.7	80	2	US-09-270-767-56655
20	7	0.7	97	2	US-09-621-976-5201
21	7	0.7	110	2	US-09-513-999C-4110
22	7	0.7	120	2	US-09-219-983A-23
23	7	0.7	120	2	US-10-114-774-23
24	7	0.7	137	2	US-09-913-204-17
25	7	0.7	162	2	US-09-270-767-41434
26	7	0.7	163	2	US-09-270-767-36487

#### ALIGNMENTS

##### RESULT 1

US-09-270-767-40274

; Sequence 40274, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 40274

; LENGTH: 179

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

; OTHER INFORMATION: Xaa means any amino acid

US-09-270-767-40274

Query Match 0.8%; Score 8; DB 2; Length 179;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAVVAS 738

Db 26 SIAAVVAS 33

##### RESULT 2

US-09-270-767-55490

; Sequence 55490, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 55490

; LENGTH: 179

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

; OTHER INFORMATION: Xaa means any amino acid

US-09-270-767-55490

Sequence 51704, A  
Sequence 14297, A  
Sequence 46614, A  
Sequence 15309, A  
Sequence 102, App  
Sequence 102, App  
Sequence 102, App  
Sequence 102, App  
Sequence 102, App  
Sequence 19, Appl  
Sequence 7649, Ap  
Sequence 7862, Ap  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 38762, A  
Sequence 53979, A  
Sequence 27564, A  
Sequence 3623, Ap

US-09-270-767-51704  
US-09-248-796A-14297  
US-09-270-767-46614  
US-09-248-796A-15309  
US-08-086-428B-102  
US-08-468-570-102  
US-08-290-665A-102  
US-08-466-601A-102  
PCT-US95-10398-102  
US-08-682-517-19  
US-09-543-681A-7649  
US-09-543-681A-7862  
US-08-208-885-2  
US-08-462-177-2  
US-08-833-622-2  
US-09-270-767-38762  
US-09-270-767-53979  
US-09-248-796A-27564  
US-09-134-000C-3623

163 2 0.7  
175 2 0.7  
186 2 0.7  
187 2 0.7  
192 1 0.7  
192 1 0.7  
192 1 0.7  
192 1 0.7  
192 2 0.7  
192 2 0.7  
192 5 0.7  
198 1 0.7  
217 2 0.7  
217 2 0.7  
217 1 0.7  
219 1 0.7  
219 1 0.7  
220 2 0.7  
220 2 0.7  
220 2 0.7  
223 2 0.7  
241 2 0.7

27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

Query Match 0.8%; Score 8; DB 2; Length 179;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAVVAS 738  
DB 26 SIAAVVAS 33  
|||||

RESULT 3  
US-08-685-992-35  
; Sequence 35, Application US/08685992  
; Patent No. 5912138  
; GENERAL INFORMATION:  
; APPLICANT: Tonks, Nicholas  
; APPLICANT: Flint, Andrew J.  
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,992  
; FILING DATE: 25-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: CSHL96-03  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 781-861-6240  
; TELEFAX: 781-861-9540  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 243 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-685-992-35

Query Match 0.8%; Score 8; DB 1; Length 243;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 VTKALASL 662  
DB 149 VTKALASL 156  
|||||

RESULT 4  
US-09-144-925-35  
; Sequence 35, Application US/09144925  
; Patent No. 5951979  
; GENERAL INFORMATION:  
; APPLICANT: Tonks, Nicholas  
; APPLICANT: Flint, Andrew J.  
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASES

; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02421-4799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/144,925  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/685,992  
; FILING DATE: July 25, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: CSHL96-03Z  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 781-861-6240  
; TELEFAX: 781-861-9540  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 243 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-144-925-35

Query Match 0.8%; Score 8; DB 1; Length 243;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 VTKALASL 662  
DB 149 VTKALASL 156  
|||||

RESULT 5  
US-09-328-352-6669  
; Sequence 669, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6669  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6669

Query Match 0.8%; Score 8; DB 2; Length 445;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 SLLDYLL 288  
DB 97 SLLDYLL 104  
|||||

RESULT 6  
US-10-332-795-7  
; Sequence 7, Application US/10332795  
; Patent No. 6946253  
; GENERAL INFORMATION:  
; APPLICANT: The University Court of the University of Glasgow  
; APPLICANT: Clements, John Barlie  
; APPLICANT: MacLean, Alasdair Roderick  
; TITLE OF INVENTION: HERPES ZINC FINGER MOTIFS  
; FILE REFERENCE: 9013-51  
; CURRENT APPLICATION NUMBER: US/10/332.795  
; CURRENT FILING DATE: 2003-01-10  
; PRIOR APPLICATION NUMBER: GB 0016890.6  
; PRIOR FILING DATE: 2000-07-11  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 7  
; LENGTH: 452  
; TYPE: PRT  
; ORGANISM: Varicella-zoster virus  
US-10-332-795-7

Query Match 0.8%; Score 8; DB 2; Length 452;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 LLENLKLK 332  
|||||  
Db 321 LLENLKLK 328

RESULT 7  
US-09-533-029-106  
; Sequence 106, Application US/09533029  
; Patent No. 666446  
; GENERAL INFORMATION:  
; APPLICANT: Heard, Jacqueline  
; APPLICANT: Broun, Pierre  
; APPLICANT: Riechmann, Jose-Luis  
; APPLICANT: Keddie, James  
; APPLICANT: Pineda, Onaira  
; APPLICANT: Adam, Luc  
; APPLICANT: Samaha, Raymond  
; APPLICANT: Zhang, James  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Pilgrim, Marsha  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Reuber, Lynne  
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES  
; FILE REFERENCE: MBI-010  
; CURRENT APPLICATION NUMBER: US/09/533.029  
; CURRENT FILING DATE: 2000-03-22  
; EARLIER APPLICATION NUMBER: 60/125,814  
; EARLIER FILING DATE: 1999-03-23  
; NUMBER OF SEQ ID NOS: 121  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 106  
; LENGTH: 969  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: G1149  
US-09-533-029-106

Query Match 0.8%; Score 8; DB 2; Length 969;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 STAAVVAS 738  
|||||  
Db 706 STAAVVAS 713

RESULT 8  
US-08-268-251-19  
; Sequence 19, Application US/08268251  
; Patent No. 5585475  
; GENERAL INFORMATION:  
; APPLICANT: Jamieson, Gordon A  
; APPLICANT: Dedman, John R  
; APPLICANT: Kaetzel, Marcia A  
; TITLE OF INVENTION: Calmodulin-Binding Peptides  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/268,251  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/831,219  
; FILING DATE: 06-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Grant D  
; REGISTRATION NUMBER: 31,259  
; REFERENCE/DOCKET NUMBER: 272.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-268-251-19

Query Match 0.7%; Score 7; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 ELRLSKS 821  
|||||  
Db 6 ELRLSKS 12

RESULT 9  
PCT-US93-01112-19  
; Sequence 19, Application PC/TUS9301112  
; GENERAL INFORMATION:  
; APPLICANT: Jamieson, Gordon A  
; APPLICANT: Dedman, John R  
; APPLICANT: Kaetzel, Marcia A  
; TITLE OF INVENTION: Calmodulin-Binding Peptides  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk



; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/01112  
; FILING DATE: 19930208  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/831,219  
; FILING DATE: 06-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Grant D  
; REGISTRATION NUMBER: 31,259  
; REFERENCE/DOCKET NUMBER: 272.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US93-01112-19

Query Match 0.7%; Score 7; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 ELRLSKS 821  
DB 6 ELRLSKS 12

RESULT 10  
US-08-086-428B-159  
; Sequence 159, Application US/08086428B  
; Patent No. 5514539  
; GENERAL INFORMATION:  
; APPLICANT: BUGH, J., MILLER, R.H. AND  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 159  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/086,428B  
; FILING DATE: 29-JUN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4070  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 159:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
US-08-086-428B-159

Query Match 0.7%; Score 7; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 SILRVPE 227  
DB 8 SILRVPE 14

RESULT 11  
US-08-268-251-49  
; Sequence 49, Application US/08268251  
; Patent No. 5585475  
; GENERAL INFORMATION:  
; APPLICANT: Jamieson, Gordon A  
; APPLICANT: Dedman, John R  
; APPLICANT: Kaetzel, Marcia A  
; TITLE OF INVENTION: Calmodulin-Binding Peptides  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/268,251  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/831,219  
; FILING DATE: 06-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Grant D  
; REGISTRATION NUMBER: 31,259  
; REFERENCE/DOCKET NUMBER: 272.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-268-251-49

Query Match 0.7%; Score 7; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 ELRLSKS 821  
DB 8 ELRLSKS 14

RESULT 12  
US-08-468-570-159  
; Sequence 159, Application US/08468570

Patent No. 5871962  
GENERAL INFORMATION:  
APPLICANT: BURKH, J., MILLER, R.H. AND  
APPLICANT: PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 159  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,570  
FILING DATE: 6-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/086,428  
FILING DATE: 29-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4070US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 159:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
US-08-468-570-159

Query Match 0.7%; Score 7; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 SILRVE 227  
Db 8 SILRVE 14

RESULT 13  
US-08-290-665A-263  
Sequence 263, Application US/08290665A  
Patent No. 5882852  
GENERAL INFORMATION:  
APPLICANT: BURKH, J., MILLER, R.H. AND  
APPLICANT: PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK

COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,665A  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 263:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
US-08-290-665A-263

Query Match 0.7%; Score 7; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 SILRVE 227  
Db 8 SILRVE 14

RESULT 14  
US-08-466-601A-159  
Sequence 159, Application US/08466601A  
Patent No. 6572864  
GENERAL INFORMATION:  
APPLICANT: BURKH, J., MILLER, R.H. AND  
APPLICANT: PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 160  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,601A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/086,428  
FILING DATE: 29-JUN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4070US2

Search completed: July 5, 2006, 14:00:40  
Job time : 54 secs

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 159:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
US-08-466-601A-159

Query Match 0.7%; Score 7; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 SILRVPE 227  
DB 8 SILRVPE 14

RESULT 15  
PCT-US93-01112-49  
Sequence 49, Application PC/TUS9301112  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzel, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01112  
FILING DATE: 19930208  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-01112-49

Query Match 0.7%; Score 7; DB 5; Length 23;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 ELRSLKS 821  
DB 8 ELRSLKS 14

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 13:50:21 ; Search time 316 Seconds  
(without alignments)  
2985.810 Million cell updates/sec

Title: US-10-645-746-3

Perfect score: 1020

Sequence: 1 MSSNPFPELEKGFYRHSLDPE.....RHEMFLOQTNVYKPGMSFA 1020

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2849598 seqs, 925015592 residues

Word size : 1

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Uniprot 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020	100.0	1020	Q9XU82_CABEL	Q9XU82 caenorhabdi
2	26	2.5	976	Q60V80_CAEBR	Q60V80 caenorhabdi
3	15	1.5	185	Q46002_CABEL	Q46002 caenorhabdi
4	11	1.1	387	Q9NXV9_HUMAN	Q9NXV9 homo sapien
5	11	1.1	794	Q5ZMW0_CHICK	Q5ZMW0 gallus gall
6	11	1.1	861	I2C4_HUMAN	Q9HCK5 homo sapien
7	11	1.1	861	Q5VXF0_HUMAN	Q5VXF0 homo sapien
8	11	1.1	884	Q4KLVE_XENLA	Q4KLVE xenopus lae
9	11	1.1	945	Q4SVE6_TETNG	Q4SVE6 tetraodon n
10	10	1.0	371	Q9NW28_HUMAN	Q9NW28 homo sapien
11	10	1.0	455	Q3TQ88_MOUSE	Q3TQ88 mus musculu
12	10	1.0	530	Q96SW6_HUMAN	Q96SW6 homo sapien
13	10	1.0	580	Q9NMV6_MOUSE	Q9NMV6 mus musculu
14	10	1.0	971	Q8CDG1_MOUSE	Q8CDG1 mus musculu
15	10	1.0	971	Q9JMB6_MOUSE	Q9JMB6 mus musculu
16	10	1.0	973	Q8TCS9_HUMAN	Q8TCS9 homo sapien
17	10	1.0	1054	Q75HC2_ORYSA	Q75HC2 oryza sativ
18	10	1.0	1121	Q61931_CABEL	Q61931 caenorhabdi
19	9	0.9	361	Q7XBH5_ORYSA	Q7XBH5 oryza sativ
20	9	0.9	414	Q9KRL3_VIBCH	Q9KRL3 vibrio chol
21	9	0.9	478	Q5ZHV0_CHICK	Q5ZHV0 gallus gall
22	9	0.9	483	Q31Q93_SYNP7	Q31Q93 synchococc
23	9	0.9	483	Q5N323_SYNP6	Q5N323 synchococc
24	9	0.9	540	Q3GBT1_9FIRM	Q3GBT1 syntrophomo
25	9	0.9	580	Q56X15_ARATH	Q56X15 arabidopsis
26	9	0.9	590	Q358Y1_9BRAD	Q358Y1 bradyrhizob
27	9	0.9	850	Q3E984_ARATH	Q3E984 arabidopsis
28	9	0.9	879	Q5NEN9_ORYSA	Q5NEN9 oryza sativ
29	9	0.9	889	Q6YJ55_ORYSA	Q6YJ55 oryza sativ
30	9	0.9	891	Q86B39_CABEL	Q86B39 caenorhabdi
31	9	0.9	892	Q84VQ0_ARATH	Q84VQ0 arabidopsis

32 9 0.9 896 2 Q84Y14\_ARATH Q84Y14 arabidopsis  
33 9 0.9 904 2 Q9SDG8\_ORYSA Q9SDG8 oryza sativ  
34 9 0.9 905 2 Q2LFC1\_NICBE Q2LFC1 nicotiana b  
35 9 0.9 910 2 Q16720\_CABEL Q16720 caenorhabdi  
36 9 0.9 912 2 Q2LFC2\_NICBE Q2LFC2 nicotiana b  
37 9 0.9 924 2 Q9ZVD5\_ARATH Q9ZVD5 arabidopsis  
38 9 0.9 948 2 Q4RPK4\_TETNG Q4RPK4 tetraodon n  
39 9 0.9 1002 2 Q20578\_CABEL Q20578 caenorhabdi  
40 9 0.9 1010 2 Q3LTR7\_CABEL Q3LTR7 caenorhabdi  
41 9 0.9 1024 2 Q61PV1\_CABER Q61PV1 caenorhabdi  
42 9 0.9 1032 2 Q7JLZ2\_CABEL Q7JLZ2 caenorhabdi  
43 9 0.9 1034 2 Q61J44\_CABER Q61J44 caenorhabdi  
44 9 0.9 1035 2 Q21079\_CABEL Q21079 caenorhabdi  
45 9 0.9 1040 1 Y043\_CABEL Y043 caenorhabdi

#### ALIGNMENTS

RESULT 1  
Q9XU82\_CABEL  
ID Q9XU82\_CABEL PRELIMINARY; PRT; 1020 AA.  
AC Q9XU82; Q9UGQ1;  
DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2001, sequence version 2.  
DT 07-FEB-2006, entry version 26.  
DE Hypothetical protein rde-1 (RNA interference promoting factor RDE- 1).  
GN Name=rde-1; ORFNames=K08H10.7;  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC SYRAIN=Bristol N2;  
RX MEDLINE=99089613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;  
RG The C. elegans sequencing consortium;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20004389; PubMed=10535731; DOI=10.1016/S0092-8674(00)81644-X;  
RA Tabara H., Sarkissian M., Kelly W.G., Pleenor J., Grishok A.,  
Timmons L., Fire A., Mello C.C.;  
RT "The rde-1 gene, RNA interference, and transposon silencing in C.  
elegans";  
RL Cell 99:123-132(1999).  
CC  
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
EMBL; Z83113; CAB05546.2; -; Genomic DNA.  
EMBL; AF180730; AAF06159.1; -; mRNA.  
PIR; T23510; T23510.  
DR Ensembl; K08H10.7; Caenorhabditis elegans.  
DR WormBase; WBGene0004323; rde-1.  
DR WormPep; K08H10.7; CE28243.  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Piwi.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Piwi; 1.  
DR PROSITE; PS0822; PIWI; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 1020 AA; 118804 MW; 8E2F1A2BFC43A670 CRC64;

Query Match 100.0%; Score 1020; DB 2; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNPFPELEKGFYRHSLDPEMKWLARPTCKDCKGFYEKKVLLVNNPKFSSKIYDREYYE 60  
DB 1 MSSNPFPELEKGFYRHSLDPEMKWLARPTCKDCKGFYEKKVLLVNNPKFSSKIYDREYYE 60

```
QY 61 YEYKMTKEVLRKPGKGFPPKKTETIPIDRAKLFWOHLRHEKKQTDPILEDYVYFDEKDTY 120
DB 61 YEYKMTKEVLRKPGKGFPPKKTETIPIDRAKLFWOHLRHEKKQTDPILEDYVYFDEKDTY 120
QY 121 SVCRLMTVTSKMLVSKVVKQSEKKDEKLEKKILYTWILTYRKFFHLNFSRENPEKDB 180
DB 121 SVCRLMTVTSKMLVSKVVKQSEKKDEKLEKKILYTWILTYRKFFHLNFSRENPEKDB 180
QY 181 EANRSYKFLKNVMTQKRVYAPFVNEBIKQVQAKNFVYDNNLSILRVSPESFHDNPNRFSQSL 240
DB 181 EANRSYKFLKNVMTQKRVYAPFVNEBIKQVQAKNFVYDNNLSILRVSPESFHDNPNRFSQSL 240
QY 241 VAPRIEAWGIYIGIKELFGEGBVLNFAVDKLFYNAPKMSLLDYLLLVDPQSCNDDYR 300
DB 241 VAPRIEAWGIYIGIKELFGEGBVLNFAVDKLFYNAPKMSLLDYLLLVDPQSCNDDYR 300
QY 301 KDLKTKLMAGKMTIROAARPRIQLLENLKLCAEAVDNEMSLRTERHLTFDLDCENSL 360
DB 301 KDLKTKLMAGKMTIROAARPRIQLLENLKLCAEAVDNEMSLRTERHLTFDLDCENSL 360
QY 361 VYKVTGSDRGNAKKYDTTLFKIYEENKKFISFPHLPLVKVKGAKYAVPMHEHLEVHE 420
DB 361 VYKVTGSDRGNAKKYDTTLFKIYEENKKFISFPHLPLVKVKGAKYAVPMHEHLEVHE 420
QY 421 KPORYKNRDLVMDQKFLKRAKPKHDYKENTLKMELDFSSSEELNFRERGLCSKLOM 480
DB 421 KPORYKNRDLVMDQKFLKRAKPKHDYKENTLKMELDFSSSEELNFRERGLCSKLOM 480
QY 481 IECPCGKVLKPEMLVNSVNEQIKMTPIRVIRFOEQKLNVPPEKELCCAVFVNNETAGNPCLE 540
DB 481 IECPCGKVLKPEMLVNSVNEQIKMTPIRVIRFOEQKLNVPPEKELCCAVFVNNETAGNPCLE 540
QY 541 ENDVVKFYTELIGCGKFRGIRIGANENRGAQSIMYDATKNEYAFYKCTLNTGIGRFETA 600
DB 541 ENDVVKFYTELIGCGKFRGIRIGANENRGAQSIMYDATKNEYAFYKCTLNTGIGRFETA 600
QY 601 ATEAKNMFRLPDKEQKVLMPFIISKROLNAGFVGHYCDHTIGVANQHITSETVTYKALA 660
DB 601 ATEAKNMFRLPDKEQKVLMPFIISKROLNAGFVGHYCDHTIGVANQHITSETVTYKALA 660
QY 661 SLRHEKSKRIFQYIAKINAKLGGINQELDSEIAEISPEEKERRKTMPLTYMGIDVT 720
DB 661 SLRHEKSKRIFQYIAKINAKLGGINQELDSEIAEISPEEKERRKTMPLTYMGIDVT 720
QY 721 HPTSYSGIDYSIAAVVASINPGGTIYRNNMIVTOEBCRPGERAVAHGRERTDILEAKFVKL 780
DB 721 HPTSYSGIDYSIAAVVASINPGGTIYRNNMIVTOEBCRPGERAVAHGRERTDILEAKFVKL 780
QY 781 LREFAENNNRAPAHITVYVRDGVSDSEMLRVSHDELSLKSEVKQFMSERDGEDPEPKYT 840
DB 781 LREFAENNNRAPAHITVYVRDGVSDSEMLRVSHDELSLKSEVKQFMSERDGEDPEPKYT 840
QY 841 FVIOQRHNTLRLRMKDKPVVNKDLTPAETDVAVAVKQWEEDMKESKETGIVNPSG 900
DB 841 FVIOQRHNTLRLRMKDKPVVNKDLTPAETDVAVAVKQWEEDMKESKETGIVNPSG 900
QY 901 TTVDKLIIVSKYKPDFFLASHHGVLTGSRPGHYTVMYDDKGMQDEYVYKMTYGLAFLSARC 960
DB 901 TTVDKLIIVSKYKPDFFLASHHGVLTGSRPGHYTVMYDDKGMQDEYVYKMTYGLAFLSARC 960
QY 961 RKPISLIPVPHVAHLSCAKELYRTYKEHYIGDYAQPRTHEMEHFLQTNVYKIPQMSFA 1020
DB 961 RKPISLIPVPHVAHLSCAKELYRTYKEHYIGDYAQPRTHEMEHFLQTNVYKIPQMSFA 1020
```

RESULT 2

```
Q60VS0_CAEBR
ID Q60VS0_CAEBR PRELIMINARY; PRT; 976 AA.
AC Q60VS0;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 7.
```

```
DE Hypothetical protein CBG19426.
GN Name=CBG19426;
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=AF16;
RX PubMed=14624247; DOI=10.1371/journal.pbio.0000045;
RA Stein L.D., Bao Z., Blasgar D., Blumenthal T., Brent M.R., Chen N.,
RA Chinwalla A., Clarke L., Clee C., Coghlan A., Coulson A.,
RA D'Eustachio P., Fitch D.H.A., Fulton L.A., Fulton R.E.,
RA Griffiths-Jones S., Harris T.W., Hillier L.W., Kamath R.,
RA Kuwabara P.E., Mardis E.R., Marra M.A., Miner T.L., Mink P.,
RA Mullikin J.C., Plumb R.W., Rogers J., Schein J.E., Sohrmann M.,
RA Spieth J., Stajich J.E., Wei C., Willey D., Wilson R.K., Durbin R.,
RA Waterston R.H.;
RT "The genome sequence of Caenorhabditis briggsae: a platform for
RT comparative genomics";
RL PLoS Biol. 1:166-192(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAC01000098; CAB72296.1; -; Genomic_DNA.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 976 AA; 114611 MW; 8B5427451C9ADCAA CRC64;

Query Match 2.5%; Score 26; DB 2; Length 976;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 HIVVYRDGVSDSEMLRVSHDELSLK 820
DB 763 HIVVYRDGVSDSEMLRVSHDELSLK 788

RESULT 3
O46002 CAEBL PRELIMINARY; PRT; 185 AA.
AC O46002;
DT 01-JUN-1998, integrated into UniProtKB/TrEMBL.
DT 01-JUN-1998, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Hypothetical protein.
GN ORFNames=ZK218.8;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; Z82085; CAB04988.1; -; Genomic_DNA.
DR PIR; T27784; T27784.
```

```

DR Ensembl; ZK218.8; Caenorhabditis elegans.
DR WormBase; WBGene00013942; ZK218.8.
DR WormPep; ZK218.8; CE16708.
DR InterPro; IPR003165; Pfam.
DR Pfam; PF02171; Pfam; 1.
DR PROSITE; PS50822; Pfam; 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 185 AA; 21081 MW; B79F7862584F11A5 CRC64;
SQ SEQUENCE 185 AA; 21081 MW; B79F7862584F11A5 CRC64;

Query Match 1.5%; Score 15; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 953 LAFLSARCRKPSLIP 967
DB 156 LAFLSARCRKPSLIP 170

RESULT 4
Q9NXV9_HUMAN PRELIMINARY; PRT; 387 AA.
AC Q9NXV9;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein FLJ20033.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC Tissue=Colon;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegaki T., Sugano S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DE EMBL; AK000040; BAA90899.1; -; mRNA.
DR Ensembl; ENSG00000134698; Homo sapiens.
DR InterPro; IPR003165; Pfam.
DR Pfam; PF02171; Pfam; 1.
DR PROSITE; PS50822; Pfam; 1.
SQ SEQUENCE 387 AA; 43196 MW; 38ABC7BDF0A585B0 CRC64;

Query Match 1.1%; Score 11; DB 2; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 83 LKINAKLGGIN 93

RESULT 5
Q5ZMW0_CHICK PRELIMINARY; PRT; 794 AA.
AC Q5ZMW0;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Hypothetical protein.
GN ORFNames=RCJMB04.1a17;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC Tissue=Brain;
MEDLINE=20450683; PubMed=10997877; DOI=10.1093/dnares/7.4.271;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process (By similarity).
CC -!- SIMILARITY: Belongs to the argonaute family.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DE EMBL; AB046787; BAB13393.1; ALU_INIT; mRNA.
DR SMR; Q9HCK3; 212-374.
DR Ensembl; ENSG00000134698; Homo sapiens.
DR HGNC; HGNC:18424; EIF2C4.
DR MIM; 607356; Gene.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.

```

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursa lymphocytes to facilitate
RT gene function analysis.";
RL Genome Biol. 6:R6-R6(2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DE EMBL; AJ719274; CAG30933.1; -; mRNA.
DR SMR; Q5ZMW0; 145-307.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; Piwi; 1.
KW Hypothetical protein.
SQ SEQUENCE 794 AA; 89531 MW; C97D19B3C096F554 CRC64;

Query Match 1.1%; Score 11; DB 2; Length 794;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 490 LKINAKLGGIN 500

RESULT 6
I2C4_HUMAN STANDARD; PRT; 861 AA.
AC Q9HCK5;
DT 14-NOV-2003, integrated into UniProtKB/Swiss-Prot.
DT 14-NOV-2003, sequence version 2.
DT 07-FEB-2006, entry version 21.
DE Eukaryotic translation initiation factor 2C 4 (eIF2C 4) (eIF-2C 4)
DE (Argonaute-4).
GN Name=EIF2C4; Synonyms=AGO4, KIAA1567;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Brain;
MEDLINE=20450683; PubMed=10997877; DOI=10.1093/dnares/7.4.271;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process (By similarity).
CC -!- SIMILARITY: Belongs to the argonaute family.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DE EMBL; AB046787; BAB13393.1; ALU_INIT; mRNA.
DR SMR; Q9HCK3; 212-374.
DR Ensembl; ENSG00000134698; Homo sapiens.
DR HGNC; HGNC:18424; EIF2C4.
DR MIM; 607356; Gene.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.

```





```

RESULT 9
Q4SV66.TETNG
AC Q4SV66.TETNG PRELIMINARY; PRT; 945 AA.
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome 21 SCAP13761, whole genome shotgun sequence.
DS ORFNames=GSTENG00012015001.
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
[1]
NUCLEOTIDE SEQUENCE.
RP PubMed=15496914; DOI=10.1038/nature03025;
RX Jallouf O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Desilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Castolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Barra G., Lardier G., Chapple C., McKernan K.J., McSwan P., Bosak S.,
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RN Nature 431:946-957(2004).
[2]
NUCLEOTIDE SEQUENCE.
RP Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DE EMBL; CAAB01013761; CAP95386.1; -; Genomic_DNA.
DR SNR; Q4SV66; 224-386.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; Piwi; 1.
SQ SEQUENCE 945 AA; 106438 MW; 9F5E76B38E8D98A1 CRC64;

Query Match 1.1%; Score 11; DB 2; Length 945;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687
DB 601 LKINAKLGGIN 611

RESULT 10
Q9NW28 HUMAN
AC Q9NW28 PRELIMINARY; PRT; 371 AA.
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 07-MAR-2006, entry version 16.
DE CDNA FLJ10351 fis, clone NT2RM2001141.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
Homo.
OX NCBI_TaxID=9606;
[1]
NUCLEOTIDE SEQUENCE.
RP PubMed=14702039; DOI=10.1038/ng1285;
RX Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa K., Fujimori K.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihara Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RN Nat. Genet. 36:40-45(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DE EMBL; AK001213; BA991558.1; -; mRNA.
DR EMBL; ENSG00000197181; Homo sapiens.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50822; Piwi; 1.
SQ SEQUENCE 371 AA; 42215 MW; 59BCF383C997D293 CRC64;

Query Match 1.0%; Score 10; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVVRGVSD 805
DB 208 IVVVRGVSD 217

RESULT 11
Q3TQ88 MOUSE
AC Q3TQ88 PRELIMINARY; PRT; 455 AA.
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE 10 days neonate medulla oblongata cDNA, RIKEN full-length enriched
DE library, clone:B830005G15 product:piwi like homolog 2 (Drosophila),
DE full insert sequence. (Fragment).
DE Names=Piwi12;
GN Mus musculus (Mouse).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

RA NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RX Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RN Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
Banaal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
Crome M.L., Della E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
Flecher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
Gustigich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
Hill D., Huminecki L., Iacono M., Ikeo K., Iwano A., Ishikawa T.,  
Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
Liu S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
Tammaja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
Grimmond S.M., Tesdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
Iida J., Inamara K., Itoh M., Katoh T., Kawaji H., Kawagashira N.,  
Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
Grimmond S., Gustigich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
Maglott D.R., Maltale L., Marchionni L., McKenzie L., Miki H.,  
Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
Santelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
Sultana R., Takenaka Y., Taylor M.S., Teddale R.D., Tomita M.,  
Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
Wilming L.G., Wyszynski B., Yanagisawa M., Yang I., Yang L.,  
Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
Sadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
Schriml L.M., Staab F., Suzuki R., Tomita M., Wagner L., Washio T.,  
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
Blake J., Boffelli D., Bojunga N., Carninci P., de Bono M.F.,  
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
Gustigich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
Sasaki H., Sato K., Schonbach C., Seya T., Shibata Y., Storch K.-F.,  
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara I., Shibata K., Itoh M.,  
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanishi M.,  
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,

RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK163647; BAB37436.1; -; mRNA.  
DR MGI; MGI:1930036; Pw112.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR InterPro; IPR003165; Pw1.  
DR Pfam; PF02171; Pw1; 1.  
DR PROSITE; PS50822; Pw1; 1.  
FT NON TER 1  
SQ SEQUENCE 455 AA; 51894 MW; 772A85A184682FCE CRC64;  
  
Query Match 1.0%; Score 10; DB 2; Length 455;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 796 IVVYRDGVSD 805  
DB 292 IVVYRDGVSD 301  
  
RESULT 12  
Q96SW6 HUMAN PRELIMINARY; PRT; 530 AA.  
AC Q96SW6;  
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
DT 01-DEC-2001, sequence version 1.  
DT 07-MAR-2006, entry version 14.  
DE CDNA FLJ14591 fis, clone NT2RM4002034, weakly similar to Homo sapiens hwi mRNA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y., Togiyasu S., Komai F., Hara R., Takeuchi K., Arita M., Inose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Koniyama M., Tashiro H., Tanigami A., Fujiwara Y., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M., Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T., Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K., Okumura K., Nagase T., Nomura N., Kikuchi H., Maehira Y., Yamashita R., Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human cDNAs,";  
RL Nat. Genet. 36:40-45 (2004).  
CC -----

CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK027497; BAB55155.1; -; mRNA.  
DR EMBL; ENSG00000197181; Homo sapiens.  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Pw1.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Pw1; 1.  
DR PROSITE; PS50821; PAZ; 1.  
DR PROSITE; PS50822; Pw1; 1.  
SQ SEQUENCE 530 AA; 60620 MW; BA4A8AADD809B32F CRC64;  
  
Query Match 1.0%; Score 10; DB 2; Length 530;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 796 IVVYRDGVSD 805  
DB 367 IVVYRDGVSD 376  
  
RESULT 13  
Q99MV6 MOUSE PRELIMINARY; PRT; 580 AA.  
AC Q99MV6;  
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2001, sequence version 1.  
DT 07-FEB-2006, entry version 18.  
DE Pw112.  
GN Name=Pw112;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Testis;  
RX MEDLINE=21175748; PubMed=11279525; DOI=10.1038/86927;  
RA Wang P.J., McCarrey J.R., Yang F., Page D.C.;  
RT "An abundance of X-linked genes expressed in spermatogonia,";  
RL Nat. Genet. 27:422-426 (2001).  
CC -----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AF285586; AAK31365.1; -; mRNA.  
DR EMBL; ENSMUSG0000003644; Mus musculus.  
DR MGI; MGI:1930036; Pw112.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Pw1.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Pw1; 1.  
DR PROSITE; PS50821; PAZ; 1.  
DR PROSITE; PS50822; Pw1; 1.  
SQ SEQUENCE 580 AA; 66550 MW; A4EA9DE8626FA136 CRC64;  
  
Query Match 1.0%; Score 10; DB 2; Length 580;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 796 IVVYRDGVSD 805  
DB 417 IVVYRDGVSD 426  
  
RESULT 14  
Q8CDG1 MOUSE PRELIMINARY; PRT; 971 AA.  
ID Q8CDG1;  
AC Q8CDG1;  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
CC -----

DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 18.  
 DE Adult male testis cDNA, RIKEN full-length enriched library,  
 DE clone:4932443D15 product:piwi like homolog 1 (Drosophila)-like, full  
 DE insert sequence.  
 GN Name=Piwi12;  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Banal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Chowdhury V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Fletcher D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.P., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keleio J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera P., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakouchi H., Ng P.,  
 RA Nilsson O., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugtara K., Sultana R., Takenaka Y., Taki K.,  
 RA Tamoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA

RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Forrest A., Gough J.,  
 RA Grimmond S., Guscinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guscinich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;

RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.",  
RL Genome Res. 10:1757-1771(2000).  
RN [8]

## NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Testis;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuchi S., Furuno M., Hanganaki T., Hara A., Hashizume W.,  
RA Hayaehida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Kurihara C., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Katsuyama T., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; AK030116; BAC26791.1; -; mRNA.  
DR Ensembl; ENSMUSG00000033644; Mus musculus.

DR MGI; MGI:1930036; Pwll2.

DR GO; GO:0005737; Cytoplasm; IDA.

DR InterPro; IPR003100; PAZ.

DR InterPro; IPR003165; Pwll.

DR Pfam; PF02170; PAZ; 1.

DR Pfam; PF02171; Pwll; 1.

DR PROSITE; PS50821; PAZ; 1.

DR PROSITE; PS50822; Pwll; 1.

SQ SEQUENCE 971 AA; 109458 MW; B63FELL1A2B9DA0F9 CRC64;

Query Match 1.0%; Score 10; DB 2; Length 971;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVSD 805

DB 808 IVVYRDGVSD 817

## RESULT 15

## QJUMB6\_MOUSE

ID QJUMB6\_MOUSE PRELIMINARY; PRT; 971 AA.

AC QJUMB6;

DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.

DT 01-OCT-2000, sequence version 1.

DT 07-FEB-2006, entry version 16.

DE MILI.

GN Name=Pwll2; Synonyms=mili;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=21463379; PubMed=11578866; DOI=10.1016/S0925-4773(01)00499-3;

RA Miyagawa S.K., Kimura T., Yomogida K., Kuroiwa A., Tadokoro Y.,

RA Fujita Y., Sato M., Matsuda Y., Nakano T.,

RT "Two mouse pwi-related genes: mili and mili."

RL Mech. Dev. 108:121-133(2001).

CC

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC

EMBL; AB032605; BAA93706.1; -; mRNA.

DR Ensembl; ENSMUSG00000033644; Mus musculus.

DR MGI; MGI:1930036; Pwll2.

DR GO; GO:0005737; Cytoplasm; IDA.

DR InterPro; IPR003100; PAZ.

DR InterPro; IPR003165; Pwll.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Pwll; 1.  
DR PROSITE; PS50821; PAZ; 1.  
DR PROSITE; PS50822; Pwll; 1.  
SQ SEQUENCE 971 AA; 109488 MW; 01E143C6513310FB CRC64;

Query Match 1.0%; Score 10; DB 2; Length 971;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVSD 805

DB 808 IVVYRDGVSD 817

Search completed: July 5, 2006, 13:58:50  
Job time : 319 secs

THIS PAGE RI ANK n1c07n1

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: July 5, 2006, 13:53:42 ; Search time 48 Seconds  
(without alignments)  
2044.606 Million cell updates/sec

Title: US-10-645-746-3

Perfect score: 1020

Sequence: 1 MSSNFPLEKGFYRHSIDP.....RHEMHPLOTNVTKTPGMSFA 1020

Scoring table: OLIGO

Searched: 283416 seqs, 96216763 residues

Word size: 1

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

PIR\_80: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	45.9	905	2 T23510	hypothetical prote
2	15	1.5	185	2 T27784	hypothetical prote
3	10	1.0	1139	2 T33275	hypothetical prote
4	9	0.9	414	2 A82177	conserved hypothet
5	9	0.9	891	2 T32079	hypothetical prote
6	9	0.9	930	2 A84668	Argonaute (AGO1)-1
7	9	0.9	958	2 S41013	hypothetical prote
8	9	0.9	1000	2 T22391	hypothetical prote
9	9	0.9	1032	2 T23164	hypothetical prote
10	9	0.9	1035	2 T23165	hypothetical prote
11	9	0.9	1040	2 D88568	protein ZK757.3 [i
12	8	0.8	192	2 S59537	heat shock transcr
13	8	0.8	452	1 WZB84	gene 4 protein - h
14	8	0.8	453	2 B70195	response regulator
15	8	0.8	468	2 A53889	protein-cytosine-p
16	8	0.8	468	2 T43622	targeted effector
17	8	0.8	468	2 S01054	variance protein
18	8	0.8	477	2 T05202	pectinesterase hom
19	8	0.8	487	2 G95388	probable aldehyde
20	8	0.8	507	2 AE0473	colicin (partial)
21	8	0.8	576	2 S65001	probable membrane
22	8	0.8	988	2 T52134	Zwille protein [m
23	8	0.8	997	2 A84678	Argonaute (AGO1)-1
24	8	0.8	1372	2 F97722	hypothetical prote
25	8	0.8	1372	2 B71724	dna-directed RNA p
26	8	0.8	1526	2 T41522	myosin ii - fibrilo
27	7	0.7	62	2 C71571	hypothetical prote
28	7	0.7	62	2 E81724	conserved hypothet
29	7	0.7	85	1 A69886	conserved hypothet

30	7	0.7	88	2 G70331	anti sigma factor
31	7	0.7	93	2 A97726	hypothetical prote
32	7	0.7	95	2 T17252	hypothetical prote
33	7	0.7	112	2 E83814	Na/H+ antiporter
34	7	0.7	112	2 S61394	hypothetical prote
35	7	0.7	115	2 H71104	hypothetical prote
36	7	0.7	123	1 HQDVFS	hydrogenase (EC 1.
37	7	0.7	130	2 T08329	hypothetical prote
38	7	0.7	139	1 W6MLB2	E6 protein - bovin
39	7	0.7	150	1 S36991	transposase (clone
40	7	0.7	158	2 W6MLPR	E6 protein - human
41	7	0.7	166	2 F96024	conserved hypothet
42	7	0.7	193	2 E95340	hypothetical prote
43	7	0.7	194	2 B97727	prolyl endopeptida
44	7	0.7	196	2 D84380	indololglycerol-
45	7	0.7	199	2 C88021	protein M10D9.1 [i

## ALIGNMENTS

## RESULT 1

T23510  
hypothetical protein K08H10.7 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T23510

R/Gardner, A.  
Submitted to the EMBL Data Library, November 1996  
A/Reference number: Z19750

A/Accession: T23510  
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-905 <WIL>

A/Cross-references: UNIPROT:Q9XU82; UNIPARC:UPI000017BAA9; EMBL:Z83113; PIDN:CAB05546.1

A/Experimental source: Clone K08H10

C/Genetics:  
A/Gene: CESP:K08H10.7

A/Map position: 5  
A/Introns: 19/3; 86/2; 190/3; 209/2; 269/1; 341/2; 468/3; 671/3; 832/3

Query Match 45.9%; Score 468; DB 2; Length 905;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSSNFPLEKGFYRHSIDPEKMLARPTGKCDGKFKYKTLILVNMFKFSKITYDREYIE	60
Db	1	MSSNFPLEKGFYRHSIDPEKMLARPTGKCDGKFKYKTLILVNMFKFSKITYDREYIE	60
Qy	61	YEVMTKEVLRNRKRGKPEPKKTEIPIDRAKLFWQHLRHEKKOTDFTLEDVVPDEKDTVY	120
Db	61	YEVMTKEVLRNRKRGKPEPKKTEIPIDRAKLFWQHLRHEKKOTDFTLEDVVPDEKDTVY	120
Qy	121	SVCLRLNTVTSKMLVSEKVVKKDSEKDEKLEKILYTWLITVYRKKFHLNFSRENPEKDE	180
Db	121	SVCLRLNTVTSKMLVSEKVVKKDSEKDEKLEKILYTWLITVYRKKFHLNFSRENPEKDE	180
Qy	181	EANRSYFELKNVNTQKRYAPFVNEEIKVOPAKNFVYDNNISILRVPSFFDPNRFESGLE	240
Db	181	EANRSYFELKNVNTQKRYAPFVNEEIKVOPAKNFVYDNNISILRVPSFFDPNRFESGLE	240
Qy	241	VAPRIEAFGIYIGIKELPDGEPVLPALVDKLFYVAPKMSLLDYLLIYDPOSCHNDVR	300
Db	241	VAPRIEAFGIYIGIKELPDGEPVLPALVDKLFYVAPKMSLLDYLLIYDPOSCHNDVR	300
Qy	301	KDLTKLMAGKMTIRQARPRIRQLLENTLAKCAEVDNEMSRUTERHLLTFDLCEBNSL	360
Db	301	KDLTKLMAGKMTIRQARPRIRQLLENTLAKCAEVDNEMSRUTERHLLTFDLCEBNSL	360
Qy	361	VYKVTGSDRGNAKKYDTDTLFTIYENKKFIFISPHLPYLVKVSAGAKYAVPMHLEVHS	420
Db	361	VYKVTGSDRGNAKKYDTDTLFTIYENKKFIFISPHLPYLVKVSAGAKYAVPMHLEVHS	420



QY 421 KFORYNRIDLVWQDKFLKRAATKPHDYKENTLKMKEIDFSSSEELNF 468  
 |||||  
 Db 421 KFORYNRIDLVWQDKFLKRAATKPHDYKENTLKMKEIDFSSSEELNF 468

## RESULT 2

hypothetical protein ZK218.8 - Caenorhabditis elegans  
 C1Species: Caenorhabditis elegans  
 C1Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C1Accession: T27784  
 R1McMurray, A.  
 submitted to the EMBL Data Library, November 1996  
 A1Reference number: Z20418  
 A1Accession: T27784  
 A1Status: preliminary; translated from GB/EMBL/DBJ  
 A1Molecule type: DNA  
 A1Residues: 1-185 <MIL>  
 A1Cross-references: UNIPROT:Q46002; UNIPARC:UPI0000079070; EMBL:Z82085; PIDN:CAE04988.1;  
 A1Experimental source: clone ZK218  
 C1Genetics:  
 A1Gene: CBSP:ZK218.8  
 A1Map position: 5  
 A1Introns: 37/3; 150/3; 174/2

Query Match 1.5%; Score 15; DB 2; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 953 LAFLSARCRKPSLP 967  
 |||||  
 Db 156 LAFLSARCRKPSLP 170

## RESULT 3

hypothetical protein R09A1.1 - Caenorhabditis elegans  
 C1Species: Caenorhabditis elegans  
 C1Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
 C1Accession: T33275  
 R1Bentley, D.; Kemp, K.; Harper, M.  
 submitted to the EMBL Data Library, May 1998  
 A1Description: The sequence of C. elegans cosmid R09A1.  
 A1Reference number: Z21312  
 A1Accession: T33275  
 A1Status: preliminary; translated from GB/EMBL/DBJ  
 A1Molecule type: DNA  
 A1Residues: 1-1139 <BEN>  
 A1Cross-references: UNIPROT:O61931; UNIPARC:UPI000017BB0B; EMBL:AF068711; PIDN:AA117775.  
 A1Experimental source: strain Bristol N2; clone R09A1  
 C1Genetics:  
 A1Gene: CBSP:R09A1.1  
 A1Map position: 5  
 A1Introns: 314/2; 376/1; 682/2; 774/1; 818/3; 1106/2

Query Match 1.0%; Score 10; DB 2; Length 1139;  
 Best Local Similarity 100.0%; Pred. No. 0.32;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 PAHIVYRDG 802  
 |||||  
 Db 940 PAHIVYRDG 949

## RESULT 4

conserved hypothetical protein VC1624 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C1Species: Vibrio cholerae  
 C1Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
 C1Accession: AB2177  
 R1Heidelberg, J.F.; Eilen, J.A.; Nelson, W.C.; Clayton, R.A.; Gaim, M.L.; Dodson, R.J.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragoi, I.; Sellers, F.  
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000  
 A1Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A1Reference number: AB2035; MID:20406833; PMID:10952301

A1Accession: AB2177  
 A1Status: preliminary  
 A1Molecule type: DNA  
 A1Residues: 1-414 <HEI>  
 A1Cross-references: UNIPROT:Q9KRL3; UNIPARC:UPI00000C3055; GB:AE004240; GB:AE003852; NIT  
 C1Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C1Genetics:  
 A1Gene: VC1624  
 A1Map position: 1  
 C1Superfamily: Synechocystis hypothetical protein slr0049

Query Match 0.9%; Score 9; DB 2; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 812 SHDELRSKL 820  
 |||||  
 Db 238 SHDELRSKL 246

## RESULT 5

hypothetical protein T07D3.7 - Caenorhabditis elegans  
 C1Species: Caenorhabditis elegans  
 C1Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C1Accession: T32079  
 R1Lamar, B.; Wamley, P.; Twyman, B.  
 submitted to the EMBL Data Library, July 1997  
 A1Description: The sequence of C. elegans cosmid T07D3.  
 A1Reference number: Z21121  
 A1Accession: T32079  
 A1Status: preliminary; translated from GB/EMBL/DBJ  
 A1Molecule type: DNA  
 A1Residues: 1-891 <LAM>  
 A1Cross-references: UNIPARC:UPI000002DB7; EMBL:AF016682; PIDN:AA66187.1; GSPDB:GNO002;  
 A1Experimental source: strain Bristol N2; clone T07D3  
 C1Genetics:  
 A1Gene: CBSP:T07D3.7  
 A1Map position: 2  
 A1Introns: 34/2; 92/2; 206/1; 560/1; 879/3  
 C1Superfamily: rabbit translation initiation factor eIF-2C

Query Match 0.9%; Score 9; DB 2; Length 891;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGV 804  
 |||||  
 Db 694 IVVYRDGV 702

## RESULT 6

Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana  
 C1Species: Arabidopsis thaliana (mouse-ear cress)  
 C1Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 C1Accession: AB4668  
 R1Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
 eues, D.; Niernman, W.C.; White, O.; Eilen, J.A.; Salberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A1Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A1Reference number: AB4420; MID:20083467; PMID:10617197  
 A1Accession: AB4668  
 A1Status: preliminary  
 A1Molecule type: DNA  
 A1Residues: 1-930 <STO>  
 A1Cross-references: UNIPROT:Q9ZVD5; UNIPARC:UPI0000179101; GB:AE002093; NID:93885334; P  
 C1Genetics:  
 A1Gene: AT2G27040

A;Map position: 2  
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 0.9%; Score 9; DB 2; Length 930;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGG 685  
|||||  
Db 631 LKINAKLGG 639

RESULT 7  
S41013  
hypothetical protein ZK757.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 31-Dec-2004  
C;Accession: S41013  
R;Thomas, K.  
submitted to the EMBL Data Library, December 1993  
A;Reference number: S41011  
A;Accession: S41013  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-958 <THO>  
A;Cross-references: UNIPROT:Q9TW94; UNIPARC:UPI0000179102; EMBL:Z29121  
C;Genetics:  
A;Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1

Query Match 0.9%; Score 9; DB 2; Length 958;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 715 VGIDVTHPT 723  
|||||  
Db 658 VGIDVTHPT 666

RESULT 8  
T22391  
hypothetical protein F48F7.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T22391; T23234  
R;Coles, L.  
submitted to the EMBL Data Library, February 1996  
A;Reference number: Z19559  
A;Accession: T22391  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1000 <WIL>  
A;Cross-references: UNIPROT:Q20578; UNIPARC:UPI0000179103; EMBL:Z69661; PIDN:CAA93496.1;  
A;Experimental source: clone F48F7  
R;Coles, L.  
submitted to the EMBL Data Library, February 1996  
A;Reference number: Z19712  
A;Accession: T22334  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1000 <WI2>  
A;Cross-references: UNIPARC:UPI0000179103; EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN000028;  
A;Experimental source: clone K02B9  
C;Genetics:  
A;Gene: CESP:F48F7.1  
A;Map position: X  
A;Introns: 70/3; 128/2; 185/2; 673/1  
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 0.9%; Score 9; DB 2; Length 1000;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IIVYRDGVS 804

A;Map position: 2  
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 0.9%; Score 9; DB 2; Length 1032;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 715 VGIDVTHPT 723  
|||||  
Db 732 VGIDVTHPT 740

RESULT 10  
T23165  
hypothetical protein T22B3.2b - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T23165; T25100  
R;Cottage, A.  
submitted to the EMBL Data Library, January 1996  
A;Reference number: Z19701  
A;Accession: T23165  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1035 <WIL>  
A;Cross-references: UNIPROT:Q21079; UNIPARC:UPI000008666E; EMBL:Z68750; PIDN:CAA92970.1  
A;Experimental source: clone K01A6  
R;Lennard, N.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: Z19981  
A;Accession: T25100  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1035 <WI2>  
A;Cross-references: UNIPARC:UPI000008666E; EMBL:Z68300; PIDN:CAA92619.1; GSPDB:GN000022;  
A;Experimental source: clone T22B3  
C;Genetics:  
A;Gene: CESP:T22B3.2b  
A;Map position: 4  
A;Introns: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 798/2; 964/2; 1006/1  
C;Superfamily: rabbit translation initiation factor eIF-2C

A;Map position: 2  
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 0.9%; Score 9; DB 2; Length 1032;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 715 VGIDVTHPT 723  
|||||  
Db 732 VGIDVTHPT 740

RESULT 10  
T23165  
hypothetical protein T22B3.2a - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T23165; T25099  
R;Cottage, A.  
submitted to the EMBL Data Library, January 1996  
A;Reference number: Z19701  
A;Accession: T23164  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1032 <WIL>  
A;Cross-references: UNIPROT:Q21079; UNIPARC:UPI000002A1C2; EMBL:Z68750; PIDN:CAA92969.1  
A;Experimental source: clone K01A6  
R;Lennard, N.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: Z19981  
A;Accession: T25099  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1032 <WI2>  
A;Cross-references: UNIPARC:UPI000002A1C2; EMBL:Z68300; PIDN:CAA92618.1; GSPDB:GN000022;  
A;Experimental source: clone T22B3  
C;Genetics:  
A;Gene: CESP:T22B3.2a  
A;Map position: 4  
A;Introns: 19/1; 71/3; 129/2; 424/2; 496/3; 607/1; 666/1; 753/3; 795/2; 961/2; 1003/1  
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 0.9%; Score 9; DB 2; Length 1035;  
Best Local Similarity 100.0%; Pred. No. 3.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0

QY 715 VGIDVTHPT 723  
|||||  
Db 735 VGIDVTHPT 743

RESULT 11  
D88568  
protein ZK757.3 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: D88568  
R:Anonymous, The C. elegans Sequencing Consortium.  
S:Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A:Status: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: D88568  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1040 <STO>  
A:Cross-references: UNIPROT:P34681; UNIPARC:UPI000013BC28; GB:chr\_III; PIDN:CAA82941.1;  
C:Genetics:  
A:Gene: ZK757.3  
A:Map position: 3  
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 0.9%; Score 9; DB 2; Length 1040;  
Best Local Similarity 100.0%; Pred. No. 3.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0

QY 715 VGIDVTHPT 723  
|||||  
Db 740 VGIDVTHPT 748

RESULT 12  
S59537  
heat shock transcription factor HSP21 - soybean (fragment)  
C:Species: Glycine max (soybean)  
C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: S59537; S52303  
R:Czarnecka-Verner, E.; Yuan, C.X.; Fox, P.C.; Gurley, W.B.  
Plant Mol. Biol. 29, 37-51, 1995  
A:Title: Isolation and characterization of six heat shock transcription factor cDNA clones  
A:Reference number: S59537; MUID:96017612; PMID:7579166  
A:Accession: S59537  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-192 <CZA>  
A:Cross-references: UNIPROT:Q43454; UNIPARC:UPI0000177D9B; EMBL:Z46952  
C:Genetics:  
A:Gene: HSP21  
C:Superfamily: tomato heat shock transcription factor HSP24; HSF DNA-binding domain homolog  
C:Keywords: DNA binding; heat shock; leucine zipper; nucleus; stress-induced protein; transcription factor  
F:46-141/Domain: HSF DNA-binding domain homology <HSP>

Query Match 0.8%; Score 8; DB 2; Length 192;  
Best Local Similarity 100.0%; Pred. No. 8.5; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0

QY 627 RQLNAYGF 634  
|||||  
Db 99 RQLNAYGF 106

RESULT 13  
WZBE4  
gene 4 protein - human herpesvirus 3

Query Match 0.8%; Score 8; DB 1; Length 452;  
Best Local Similarity 100.0%; Pred. No. 18; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0

QY 325 LLENLKLK 332  
|||||  
Db 321 LLENLKLK 328

RESULT 14  
B70195  
response regulatory protein (rrp-2) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 02-Jun-2003  
C:Accession: B70195  
R:Praser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whiston, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugl; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horat, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: B70195  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-453 <KLE>  
A:Cross-references: UNIPARC:UPI00000575DE; GB:AE001176; GB:AE000783; MUID:g2688699; PIDN:G2688699  
A:Experimental source: strain B31  
C:Superfamily: response regulator, NtrC type; response regulator homology; RNA polymerase  
C:Keywords: phosphoprotein  
F:6-115/Domain: response regulator homology <RRH>  
F:147-369/Domain: RNA polymerase sigma factor interaction domain homology <SFI>  
F:54/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 2; Length 453;  
Best Local Similarity 100.0%; Pred. No. 18; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0

QY 694 BIAEISPE 701  
|||||  
Db 247 BIAEISPE 254

RESULT 15  
A53889  
protein-tyrosine-phosphatase (EC 3.1.3.48) Yop51 - Yersinia enterocolitica (strain W227)  
N:Alternate names: virulence protein Yop51  
C:Species: Yersinia enterocolitica  
C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 09-Jul-2004  
C:Accession: A53889  
R:Michiels, T.; Cornelis, G.  
Microb. Pathog. 5, 449-459, 1988  
A:Title: Nucleotide sequence and transcription analysis of yop51 from Yersinia enterocolitica  
A:Reference number: A53889; MUID:89218619; PMID:3244311  
A:Accession: A53889  
A:Status: preliminary  
A:Molecule type: DNA

A;Residues: 1-468 <MIC>  
A;Cross-references: UNIPROT:P15273; UNIPARC:UPI00000000256; GB:M30457; NID:g155530; PIDN:  
C;Genetics:  
A;Gene: yop51  
A;Genome: plasmid  
C;Keywords: phosphoric monoester hydrolase; tyrosine-specific phosphatase  
Query Match 0.8%; Score 8; DB 2; Length 468;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 655 VTKALASL 662  
| | | | |  
Db 364 VTKALASL 371  
Search completed: July 5, 2006, 13:59:43  
Job time : 50 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM protein - protein search, using sw model

Run on: July 5, 2006, 13:49:57 ; Search time 196 Seconds  
(without alignments)  
2379.392 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 1020  
Sequence: 1 MSNPFLEKGFVHSLDPE.....RHMEHFLQTNVYKPGMSFA 1020

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2589679 seqs, 457216429 residues

Word size : 1

Total number of hits satisfying chosen parameters: 2589342

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq.8.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*  
9: Geneseqp2005s.\*  
10: Geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	100.0	1020	4	Aau01856 C. elegans
2	11	1.1	861	10	Aef73792 Human Arg
3	11	1.1	924	7	Adj69189 Human hea
4	11	1.1	924	8	Adj94699 Human euk
5	10	1.0	371	4	Aab27222 Human pro
6	10	1.0	530	4	Aab94209 Human pro
7	10	1.0	530	7	Adj69986 Human hea
8	10	1.0	580	4	Aau07866 Polypepti
9	10	1.0	678	8	Adx66443 Plant ful
10	9	0.9	94	10	Aef73781 Caenorhab
11	9	0.9	94	10	Aef73783 Caenorhab
12	9	0.9	94	10	Aef73782 Caenorhab
13	9	0.9	389	6	Adb17506 Wheat pos
14	9	0.9	389	9	Aec75730 Wheat Arg
15	9	0.9	850	8	Adq80691 Arabidops
16	9	0.9	892	8	Adq80689 Arabidops
17	9	0.9	901	6	Adb17464 Corn post
18	9	0.9	901	9	Aec75688 Corn post
19	9	0.9	904	6	Adb17511 Rice post
20	9	0.9	904	9	Aec75735 Rice post
21	9	0.9	915	6	Adb17496 Soybean p
22	9	0.9	915	9	Aec75720 Soybean p
23	9	0.9	916	6	Adb17498 Wheat pos

24	9	0.9	916	9	Aec75722 Wheat Arg
25	9	0.9	924	8	Adq80687 Arabidops
26	9	0.9	930	7	Adq31435 Plant yie
27	9	0.9	930	8	Adi44099 Plant tra
28	8	0.8	47	3	Aag50116 Arabidops
29	8	0.8	108	10	Aef73786 Arabidops
30	8	0.8	108	10	Aef73787 Arabidops
31	8	0.8	108	10	Aef73784 Rice Argo
32	8	0.8	108	10	Aef73785 Rice Argo
33	8	0.8	109	4	Aau53734 Propionib
34	8	0.8	109	6	Abm50253 Propionib
35	8	0.8	121	6	Adm06608 Alloiococ
36	8	0.8	126	7	Adk52503 Penicilli
37	8	0.8	161	8	Adri15803 Kinase ly
38	8	0.8	183	10	Aef14816 RNA-depen
39	8	0.8	235	8	Adr28077 NPB polyp
40	8	0.8	243	4	Aab59398 Yarsinia
41	8	0.8	248	8	Ady07084 Plant ful
42	8	0.8	287	4	Aag78296 Yarsinia
43	8	0.8	306	10	Aef14819 RNA-depen
44	8	0.8	326	3	Aag52610 Arabidops
45	8	0.8	337	8	Adx77635 Plant ful

#### ALIGNMENTS

RESULT 1  
AAU01856  
ID AAU01856 standard; protein; 1020 AA.  
XX  
AC AAU01856;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE C. elegans RNA interference pathway protein RDE-1.  
XX  
KW RNA interference; RNAi; RDE-1; genetic interference; antibody; dsRNA;  
KW double-stranded RNA-dependent gene silencing.  
XX  
OS Caenorhabditis elegans.  
XX  
PN WO200129058-A1.  
XX  
PD 26-APR-2001.  
XX  
PF 13-OCT-2000; 2000WO-US028470.  
XX  
PR 15-OCT-1999; 99US-0159776P.  
PR 30-MAR-2000; 2000US-0193218P.  
XX  
(UYMA-) UNIV MASSACHUSETTS.  
(CARN-) CARNEGIE INST WASHINGTON.  
XX  
Mello CC, Fire A, Tabara H, Grishok A;  
WPI; 2001-316239/33.  
DR N-PSDB; AAS03282, AAS03283.  
XX  
Novel RNA interference pathway genes and their protein products involved in mediation of genetic interference, useful for modulating and studying regulation of RNA interference pathway.  
XX  
Claim 3; Fig 6; 76pp; English.

CC The sequence represents the RNA interference (RNAi) pathway protein RDE-1. RDE-1 and RDE-4 are involved in the pathway mediating double-stranded RNA-dependent gene silencing (genetic interference) RDE-1 and RDE-4 protein is useful for preparing an RNAi agent, by incubating a dsRNA in the presence of the proteins. The prepared RNAi agents can be used as sequence-specific interfering agents for targeted genetic interference. CC The nucleic acids are useful for studying the regulation of RNAi pathway and to generate knockout strains of animals such as C.elegans. RDE-1 and

CC	RDE-4 genes and their products are useful for modulating RNAi pathway activity. The polypeptides are useful for generating and testing	
CC	antibodies specific for the polypeptides which are useful for studying	
CC	the RNAi pathway in C.elegans and other organisms. RNAi pathway genes are	
CC	useful for mediating specific processes, e.g. a gene that mediates dsRNA	
CC	uptake by the cells is useful for transporting other RNAs into cells or	
CC	for facilitating entry of agents such as drugs into cells. RNAi pathway	
CC	mutant strains (rde-1, rde-4 mutants) are useful in genetic screens to	
CC	identify additional RNAi pathway components	
XX	Sequence 1020 AA;	
Qy	Query Match 100.0%; Score 1020; DB 4; Length 1020;	
Db	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 MSSNPPELEKGFYRHSLDPEMKWLPARTGKCDGKPFYEKKVLLVNMWPKFSSKIYDREYYE 60	
Db	1 MSSNPPELEKGFYRHSLDPEMKWLPARTGKCDGKPFYEKKVLLVNMWPKFSSKIYDREYYE 60	
Qy	61 YEVMKTEVNLNRPKPGPKKTEIPIDRAKLFQWHLRHEKKQTDPILEDYVDEKDTVY 120	
Db	61 YEVMKTEVNLNRPKPGPKKTEIPIDRAKLFQWHLRHEKKQTDPILEDYVDEKDTVY 120	
Qy	121 SVCRLNTVTSKMLVSEKVKVDSEKKDEKLEKKILYTWILTYRKKFHLNFSRENPEKDE 180	
Db	121 SVCRLNTVTSKMLVSEKVKVDSEKKDEKLEKKILYTWILTYRKKFHLNFSRENPEKDE 180	
Qy	181 EANRSYKPLQNVMTQKRVYAPFVNEEIKVQAKNFVYDNNLSILRVPSFHPDRPEQSLE 240	
Db	181 EANRSYKPLQNVMTQKRVYAPFVNEEIKVQAKNFVYDNNLSILRVPSFHPDRPEQSLE 240	
Qy	241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLILLVDPQSCNDDVR 300	
Db	241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLILLVDPQSCNDDVR 300	
Qy	301 KDKTKLMAGKMTIROAAPRIQLLENLKLCAEVDNEMSLTERHLTFDLCEENSL 360	
Db	301 KDKTKLMAGKMTIROAAPRIQLLENLKLCAEVDNEMSLTERHLTFDLCEENSL 360	
Qy	361 VYKVTGSDRGNAKYYDTTLFKIYEENKFTIEFPHLPLVKVSKAKYAVAPMEHLEVHE 420	
Db	361 VYKVTGSDRGNAKYYDTTLFKIYEENKFTIEFPHLPLVKVSKAKYAVAPMEHLEVHE 420	
Qy	421 KPORYKNRIJLVMDQKFLKRAFKDHYKENTLKMELDPSEELNPFVERGLCKLOM 480	
Db	421 KPORYKNRIJLVMDQKFLKRAFKDHYKENTLKMELDPSEELNPFVERGLCKLOM 480	
Qy	481 IECPGKVLKEPMLVNSVNEQIKMTPVIRGFQSKQLNVVPEKELCCAVFVNNETAGNPCLE 540	
Db	481 IECPGKVLKEPMLVNSVNEQIKMTPVIRGFQSKQLNVVPEKELCCAVFVNNETAGNPCLE 540	
Qy	541 ENDVVKFYTELIGCGCKFRGIRGANENRGAQSIMYDATQNEYAFYKNCNTLNTGIGRFEIA 600	
Db	541 ENDVVKFYTELIGCGCKFRGIRGANENRGAQSIMYDATQNEYAFYKNCNTLNTGIGRFEIA 600	
Qy	601 ATEAKNMPERLPDKEQKLMFIIISKRLNAGFVKYCHDHTIGVANQHITSETVTKALA 660	
Db	601 ATEAKNMPERLPDKEQKLMFIIISKRLNAGFVKYCHDHTIGVANQHITSETVTKALA 660	
Qy	661 SLRHEGSKRIFYQIALKINAKLGGINQBLDSEIAEISPEEKERRKTMPLTWYVGIDVT 720	
Db	661 SLRHEGSKRIFYQIALKINAKLGGINQBLDSEIAEISPEEKERRKTMPLTWYVGIDVT 720	
Qy	721 HPTYSYGIDYSIAAVVASINPGGTIYRNMIVTQECRPGERAVAHGRERTDILEAKFVKL 780	
Db	721 HPTYSYGIDYSIAAVVASINPGGTIYRNMIVTQECRPGERAVAHGRERTDILEAKFVKL 780	
Qy	781 LREFAENNDRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKVKOPMSRDEGDEPEPKYT 840	
Db	781 LREFAENNDRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKVKOPMSRDEGDEPEPKYT 840	
Qy	841 FVIOQRNTRLRRMEKDKPVVNKDLTPAETDVAVAVKQWEDMKESKETGIYNPSSG 900	

Db	841 FVIOQRNTRLRRMEKDKPVVNKDLTPAETDVAVAVKQWEDMKESKETGIYNPSSG 900	
Qy	901 TTVDKLIIVSKYKDFFLASHHGVLTGTSRPGHVTVMYDDKGMQSDQDEVYKMTYGLAFLSARC 960	
Db	901 TTVDKLIIVSKYKDFFLASHHGVLTGTSRPGHVTVMYDDKGMQSDQDEVYKMTYGLAFLSARC 960	
Qy	961 RKPISLPVPVHYAHLSCSKAKELYRTEHYHIGDYAQRTRHEMSEHFLQTNVVKYPGMSFA 1020	
Db	961 RKPISLPVPVHYAHLSCSKAKELYRTEHYHIGDYAQRTRHEMSEHFLQTNVVKYPGMSFA 1020	
RESULT 2		
AEF73792		
ID	AEF73792 standard; protein; 861 AA.	
XX		
AC	AEF73792;	
DT	06-APR-2006 (first entry)	
XX		
DE	Human Argonaute protein hAgo4.	
XX		
KW	Argonaute; hAgo4; RNA interference; RNAi; drug screening;	
KW	X-ray crystallography; gene silencing; protein structure.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	222..342
FT		/label = PAZ_domain
FT	Domain	575..834
FT		/label = PIWI_domain
XX		
XX	WO2006015258-A2.	
PN		
XX	09-FEB-2006.	
PD		
XX	28-JUL-2005; 2005WO-US027084.	
PF		
XX		
PR	28-JUL-2004; 2004US-0592297P.	
PR	29-JUL-2004; 2004US-0592269P.	
XX		
PA	(COLD-) COLD SPRING HARBOR LAB.	
XX		
PI	Joshua-Tor L, Song J, Hannon GJ, Liu J, Carmell MA, Rivas F;	
PI	Marsden C;	
XX		
DR	WPI; 2006-155768/16.	
XX		
PT	Crystalline Argonaute useful for identifying agent that binds Argonaute	
PT	protein, utilized for treating pancreatic cancer, treating hepatitis C	
PT	infection, cancer or inflammatory diseases.	
XX		
PS	Disclosure; SEQ ID NO 4; 215pp; English.	
XX		
CC	The present sequence is that of human Argonaute protein hAgo4. Argonaute	
CC	proteins are involved in RNA interference (RNAi). The invention provides	
CC	a crystalline Argonaute protein, which comprises (i) an N-terminus,	
CC	middle and PIWI domain which form a crescent-shaped base, and (ii) a PAZ	
CC	domain, which is positioned above the crescent-shaped base, resulting in	
CC	a cleft between the crescent-shaped base and the PAZ domain. The	
CC	structure of the full-length Argonaute protein AEF73751 from the	
CC	archaeobacterium Pyrococcus furiosus (Pfi), as determined by x-ray	
CC	crystallography to 2.25 Angstrom resolution, is provided. The invention	
CC	also provides an isolated complex comprising an Argonaute protein and a	
CC	single-stranded RNA hybridized to its target nucleic acid. In certain	
CC	embodiments, the single-stranded RNA is bound to the PAZ domain of the	
CC	Argonaute protein, and may further interact with the crescent-shaped base	
CC	of the Argonaute protein. Also claimed are: a method of determining the	
CC	three-dimensional structure of an Argonaute protein or its mutant,	
CC	derivative, variant, analog, homolog, sub-domain or fragment by alignment	
CC	with the Pfi amino acid sequence to match homologous regions; a method	
CC	of identifying an agent that binds an Argonaute protein by applying a	



three-dimensional molecular modeling algorithm to the atomic coordinates of an Argonaute protein to determine the spatial coordinates of the binding pocket of the Argonaute protein, and electronically screening the stored spatial coordinates of a set of candidate agents against the spatial coordinates of the Argonaute protein binding pocket; a computer-based method for the analysis of the interaction of a molecular structure with an Argonaute protein; a data array comprising the atomic coordinates of an Argonaute protein; a computer-readable storage medium encoded with the atomic coordinates of an Argonaute protein; an electronic representation of a domain or binding site of the Argonaute protein; a method for obtaining a crystal of an Argonaute protein; a method for identifying an agent that modulates the activity of an RNAi construct by identifying an agent that modulates the expression and/or activity of an Argonaute protein; a method for identifying an agent that potentiates the activity of an RNAi construct by identifying an agent that increases the expression and/or activity of an Argonaute protein; a method of identifying an agent that modulates the activity of an RNAi construct by examining a change in Argonaute protein activity in the presence of a candidate agent; a composition for targeted gene inhibition comprising an agent that modulates the RNase activity of an Argonaute protein; a cell line that overexpresses an Argonaute protein; an assay for identifying nucleic acid sequences for conferring a particular phenotype in a cell; a nucleic acid composition composed of a nucleic acid comprising an RNAi construct and a nucleic acid encoding an Argonaute protein, where the RNAi construct comprises a nucleic sequence encoding a single-strand short interfering RNA (siRNA); a pharmaceutical composition comprising the nucleic acid composition; and a cell expressing the nucleic acid composition. The methods and compositions of the invention are useful for enhancing the effectiveness of an RNAi therapeutic.

XX SQ Sequence 861 AA;

Query Match 1.1%; Score 11; DB 10; Length 861;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 677 LKINAKLGGIN 687  
Db 557 LKINAKLGGIN 567  
|||||

RESULT 3

ADJ69189  
ID ADJ69189 standard; protein; 924 AA.

AC ADJ69189;

XX 06-MAY-2004 (first entry)

XX Human heat mitochondrial protein as a therapeutic target SeqID995.

XX mitochondrial; human; screening assay; diabetes mellitus;  
KW Huntington's disease; osteoarthritis;  
KW Leber's hereditary optic neuropathy; LHON;  
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;  
KW osteopathic; ophthalmological; cytostatic.

XX Homo sapiens.

XX WO2003087768-A2.

XX 23-OCT-2003.

XX 04-APR-2003; 2003WO-US010870.

XX 12-APR-2002; 2002US-0372843P.

XX 17-JUN-2002; 2002US-0389987P.

XX 20-SEP-2002; 2002US-0412418P.

XX (MITO-) MITOKOR.

XX (BUCK-) BUCK INST AGE RES.

XX

PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;  
PI Warnock DE;

XX WPI; 2003-845369/78.

XX Identifying a mitochondrial target for drug screening assays and for  
PT treating diseases associated with altered mitochondrial function,  
PT comprises detecting a modified polypeptide in a sample and correlating  
PT with the disease.

XX Claim 1; SEQ ID NO 995; 180pp; English.

XX This invention relates to novel mitochondrial targets that can be used  
CC for therapeutic intervention in treating a disease associated with  
CC altered mitochondrial function. Specifically, it refers to a method for  
CC identifying proteins of the human heart mitochondrial proteome that are  
CC useful for drug screening assays, as well as therapeutic targets. The  
CC present invention describes a method for identifying such proteins that  
CC can be used in the treatment of various diseases associated with altered  
CC mitochondrial function including diabetes mellitus, Huntington's disease,  
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, nontropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytostatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.

XX SQ Sequence 924 AA;

Query Match 1.1%; Score 11; DB 7; Length 924;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 677 LKINAKLGGIN 687  
Db 620 LKINAKLGGIN 630  
|||||

RESULT 4

ADJ94699

ID ADJ94699 standard; protein; 924 AA.

AC ADJ94699;

XX 06-MAY-2004 (first entry)

XX Human eukaryotic translation initiation factor 2C3 (eIF2C3) protein.

XX cytostatic; antiinflammatory; virucide; immunosuppressive; tumour;  
KW inflammatory; infectious disease; viral infection; degenerative;  
KW autoimmune; gene therapy; Argonaute family;  
KW eukaryotic translation initiation factor 2C3; eIF2C3; human.

XX Homo sapiens.

XX WO2004007718-A2.

XX 22-JAN-2004.

XX 10-JUL-2003; 2003WO-EP007516.

XX 10-JUL-2002; 2002EP-00015532.

XX 23-AUG-2002; 2002EP-00018906.

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Tuschl T, Martinez J, Patkaniowska A, Urlaub H, Luehrmann R;

XX WPI; 2004-122948/12.

XX N-PSDB; ADJ94705.

PT New single-stranded RNA molecule having a length from 14-50 nucleotides,  
 PT useful for preventing or treating tumor, inflammatory, infectious, e.g.  
 PT viral infections, degenerative and autoimmune diseases.

XX Example; Fig 14; 73pp; English.

PS The invention relates to a novel single-stranded RNA molecule having a  
 CC length from 14-50 nucleotides where at least 14-20 of the 5'-most  
 CC nucleotides are substantially complementary to a target transcript. The  
 CC RNA molecule of the invention demonstrates cytostatic, antiinflammatory,  
 CC virucide and immunosuppressive activities and may be useful for  
 CC inhibiting the expression of a target gene in vitro or in vivo.  
 CC preferably for preventing or treating diseases associated with the  
 CC overexpression of at least one target transcript. The diseases may be  
 CC selected from tumour diseases, inflammatory diseases, infectious diseases  
 CC such as viral infections, degenerative diseases and autoimmune diseases.  
 CC Furthermore, the molecules of the invention may be utilised during gene  
 CC therapy. The current sequence is that of the human eukaryotic translation  
 CC initiation factor 2C3 (eIF2C3) protein of the invention.

XX Sequence 924 AA;

Query Match 1.1%; Score 11; DB 8; Length 924;  
 Best Local Similarity 100.0%; Pred. No. 0.24;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 LKINAKLGGIN 687  
 |||||  
 DB 620 LKINAKLGGIN 630

RESULT 5  
 AAB92722  
 ID AAB92722 standard; protein; 371 AA.

AC AAB92722;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:11144.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 11144; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-

CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention

XX Sequence 371 AA;

Query Match 1.0%; Score 10; DB 4; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVSD 805  
 |||||

DB 208 IVVYRDGVSD 217

RESULT 6  
 AAB94209  
 ID AAB94209 standard; protein; 530 AA.

AC AAB94209;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:14559.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 14559; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-

CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 530 AA;

Query Match 1.0%; Score 10; DB 4; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVSD 805  
 |||||  
 Db 367 IVVYRDGVSD 376

RESULT 7  
 ADJ69986  
 ID ADJ69986 standard; protein; 530 AA.  
 AC ADJ69986;  
 XX  
 XX 06-MAY-2004 (first entry)  
 DT  
 DE Human heat mitochondrial protein as a therapeutic target SeqID1792.  
 XX  
 XX mitochondrial; human; screening assay; diabetes mellitus;  
 KW Huntington's disease; osteoarthritis;  
 KW Leber's hereditary optic neuropathy; LHON;  
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
 KW neuroprotective; neurotropic; antidiabetic; anticonvulsant; antiarthritic;  
 KW osteopathic; ophthalmological; cytostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2003087768-A2.  
 PN  
 XX 23-OCT-2003.  
 PD  
 XX 04-APR-2003; 2003WO-US010870.  
 PF  
 XX 12-APR-2002; 2002US-0372843P.  
 PR  
 XX 17-JUN-2002; 2002US-0389987P.  
 PR  
 XX 20-SEP-2002; 2002US-0412418P.  
 XX  
 XX (MITO-) MITOKOR.  
 PA (BUCK-) BUCK INST AGE RES.  
 PA  
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;  
 PI Warnock DE;  
 PI  
 XX WPI; 2003-845369/78.  
 DR  
 XX Identifying a mitochondrial target for drug screening assays and for  
 XX treating diseases associated with altered mitochondrial function,  
 PT

PT comprises detecting a modified polypeptide in a sample and correlating  
 PT with the disease.  
 XX

PS Claim 1; SEQ ID NO 1792; 180pp; English.

XX  
 CC This invention relates to novel mitochondrial targets that can be used  
 CC for therapeutic intervention in treating a disease associated with  
 CC altered mitochondrial function. Specifically, it refers to a method for  
 CC identifying proteins of the human heart mitochondrial proteome that are  
 CC useful for drug screening assays, as well as therapeutic targets. The  
 CC present invention describes a method for identifying such proteins that  
 CC can be used in the treatment of various diseases associated with altered  
 CC mitochondrial function including diabetes mellitus, Huntington's disease,  
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
 CC compositions have neuroprotective, neurotropic, antidiabetic,  
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
 CC cytostatic activities. This polypeptide sequence is a human heart  
 CC mitochondrial protein of the invention.  
 XX

SQ Sequence 530 AA;

Query Match 1.0%; Score 10; DB 7; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVSD 805  
 |||||  
 Db 367 IVVYRDGVSD 376

RESULT 8  
 AAU07866  
 ID AAU07866 standard; protein; 580 AA.  
 XX  
 AC AAU07866;  
 XX  
 XX 18-DEC-2001 (first entry)  
 DT  
 DE Polypeptide sequence for mammalian Spgl6.  
 XX  
 XX Mammalian; reproductive-specific protein; male infertility;  
 KW spermatogenesis; sperm count disorder; anti infertility; reproduction.  
 KW Mammalia.  
 OS  
 XX WO200166752-A2.  
 PN  
 XX 13-SEP-2001.  
 PD  
 XX 07-MAR-2001; 2001WO-US007371.  
 PF  
 XX 07-MAR-2000; 2000US-0187518P.  
 PR  
 XX 12-JAN-2001; 2001US-0261557P.  
 XX  
 XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
 PA  
 XX Wang PJ, Page DC;  
 PI  
 XX WPI; 2001-570774/64.  
 DR  
 XX N-PSDB; AAS13630.  
 XX  
 XX Novel reproduction-specific protein, useful for treating disorders of  
 PT reduced sperm count, enhancing/increasing sperm count and/or sperm  
 PT activity.  
 XX  
 XX Claim 22; Fig 16; 151pp; English.  
 XX  
 CC The present invention relates to the isolation of novel mammalian and  
 CC human reproductive-specific proteins (AAU07859-AAU07899), and the nucleic  
 CC acids encoding them. The nucleic acids encoding reproductive-specific  
 CC proteins are useful for diagnosing infertility which is a result of

CC reduced sperm count, reduced sperm motility, malformed sperm or  
 CC combinations of these. The sequences of the invention are useful as  
 CC markers for spermatogonial cells, for identifying genes or proteins  
 CC characteristic of male infertility, diagnosing or aiding in the diagnosis  
 CC of infertility in men, and for contraception in which sperm production or  
 CC sperm count is reduced or defective sperm is produced. Antibodies to  
 CC reproductive-specific proteins are useful for determining the presence of  
 CC these proteins in a sample obtained from a man being assessed for  
 CC infertility, for identifying the expression of genes in particular cell  
 CC type or particular developmental stage, for studies of spermatogenesis,  
 CC and for immunofluorescence of germ cells or in Western blots for  
 CC assessing the presence of the protein the antibody binds. The sequences  
 CC of the invention are also useful for treating disorders of reduced sperm  
 CC count, and for increasing sperm count and/or sperm activity. The nucleic  
 CC acids of the invention are useful in gene therapy. AAU07859-AAU07882  
 CC represent the mammalian reproduction-specific proteins of the present  
 CC invention  
 XX  
 SQ Sequence 580 AA;

Query Match 1.0%; Score 10; DB 4; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVSD 805  
 DB 417 IVVYRDGVSD 426  
 |||||

RESULT 9  
 ADX66443  
 ID ADX66443 standard; protein; 678 AA.  
 XX  
 AC ADX66443;

21-APR-2005 (first entry)

Plant full length insert polypeptide seqid 37286.

plant protectant; plant growth regulant; gene therapy; plant;  
 recombinant DNA construct; physical array; plant breeding marker;  
 cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
 extreme osmotic condition; pathogen tolerance; pest tolerance;  
 growth rate; cell cycle pathway; disease resistance;  
 Galactomannan production; lignin production; plant growth regulator;  
 yield; plant growth; plant development; seed oil; protein yield;  
 protein content.

Unidentified.

US2004034888-A1.

19-FEB-2004.

28-APR-2003; 2003US-00425114.

06-MAY-1999; 99US-00304517.

05-NOV-2001; 2001US-00985678.

(LIU//) LIU J.  
 (ZHOU//) ZHOU Y.  
 (KOVA//) KOVALIC D K.  
 (SCRE//) SCREEN S E.  
 (TAB//) TABASKA J E.  
 (CAOY//) CAO Y.

Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

WPI; 2004-180133/17.

New recombinant DNA construct, useful for improving plant tolerance to  
 cold, drought, herbicides, extreme osmotic conditions, pathogens or  
 pests, for conferring increased resistance to plant disease, or for

PT improving yield.  
 XX Claim 1; SEQ ID NO 37286; 15pp; English.  
 XX  
 CC The invention describes a recombinant DNA construct comprising a  
 CC polynucleotide consisting of a sequence encoding an amino acid sequence  
 CC available in electronic form from the US patent office at  
 CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
 CC of the invention are also useful in physical arrays of molecules and as  
 CC plant breeding markers. The recombinant DNA construct is useful for  
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
 CC plant cells by modification of the cell cycle pathway, for conferring  
 CC increased resistance to plant disease, for producing galactomannan,  
 CC lignin or plant growth regulators, for increasing the rate of homologous  
 CC recombination in plants, for improving yield by modification of  
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC or by providing improved plant growth and development under at least one  
 CC stress condition or for modifying seed oil or protein yield and/or  
 CC content. This is the amino acid sequence of a plant full length insert  
 CC polypeptide that can be used in the recombinant DNA construct of the  
 CC invention.  
 XX  
 SQ Sequence 678 AA;

Query Match 1.0%; Score 10; DB 8; Length 678;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 SIAAVVASIN 740  
 |||||

DB 440 SIAAVVASIN 449  
 |||||

RESULT 10

AEF73781

ID AEF73781 standard; protein; 94 AA.

XX AEF73781;

06-APR-2006 (first entry)

Caenorhabditis briggsae Argonaute protein Piwi domain polypeptide.

Argonaute; RNA interference; RNAi; drug screening; X-ray crystallography;  
 gene silencing.

Caenorhabditis briggsae.

WO2006015258-A2.

09-FEB-2006.

28-JUL-2005; 2005WO-US027084.

28-JUL-2004; 2004US-0592297P.

29-JUL-2004; 2004US-0592269P.

(COLD-) COLD SPRING HARBOR LAB.

Joshua-Tor L, Song J, Hannon GJ, Liu J, Carmell MA, Rivas F;  
 Marsden C;

WPI; 2006-155768/16.

Crystalline Argonaute useful for identifying agent that binds Argonaute  
 protein, utilized for treating pancreatic cancer, treating hepatitis C  
 infection, cancer or inflammatory diseases.

Example 10; Fig 21; 215pp; English.

The present invention relates to Argonaute proteins, which are involved  
 in RNA interference (RNAi). The invention provides a crystalline

CC Argonaute protein, which comprises (i) an N-terminus, middle and PIWI  
 CC domain which form a crescent-shaped base, and (ii) a PAZ domain, which is  
 CC positioned above the crescent-shaped base, resulting in a cleft between  
 CC the crescent-shaped base and the PAZ domain. The structure of the full-  
 CC length Argonaute protein AEF73751 from the archaeobacterium Pyrococcus  
 CC furiosus (PfAgol), as determined by x-ray crystallography to 2.25 Angstrom  
 CC resolution, is provided. The invention also provides an isolated complex  
 CC comprising an Argonaute protein and a single-stranded RNA hybridized to  
 CC its target nucleic acid. In certain embodiments, the single-stranded RNA  
 CC is bound to the PAZ domain of the Argonaute protein, and may further  
 CC interact with the crescent-shaped base of the Argonaute protein. Also  
 CC claimed are: a method of determining the three-dimensional structure of  
 CC an Argonaute protein or its mutant, derivative, variant, analog, homolog,  
 CC sub-domain or fragment by alignment with the PfAgol amino acid sequence to  
 CC match homologous regions; a method of identifying an agent that modulates the  
 CC Argonaute protein by applying a three-dimensional molecular modeling  
 CC algorithm to the atomic coordinates of an Argonaute protein to determine  
 CC the spatial coordinates of the binding pocket of the Argonaute protein,  
 CC and electronically screening the stored spatial coordinates of a set of  
 CC candidate agents against the spatial coordinates of the Argonaute protein  
 CC binding pocket; a computer-based method for the analysis of the  
 CC interaction of a molecular structure with an Argonaute protein; a data  
 CC array comprising the atomic coordinates of an Argonaute protein; a  
 CC computer-readable storage medium encoded with an Argonaute protein; a  
 CC site of the Argonaute protein; an electronic representation of a domain or binding  
 CC site of the Argonaute protein; a method for obtaining a crystal of an  
 CC Argonaute protein; a method for identifying an agent that modulates the  
 CC activity of an RNAi construct by identifying an agent that modulates the  
 CC expression and/or activity of an Argonaute protein; a method for  
 CC identifying an agent that potentiates the activity of an RNAi construct  
 CC by identifying an agent that increases the expression and/or activity of  
 CC an Argonaute protein; a method of identifying an agent that modulates the  
 CC activity of an RNAi construct by examining a change in Argonaute protein  
 CC activity in the presence of a candidate agent; a composition for targeted  
 CC gene inhibition comprising an agent that modulates the RNase activity of  
 CC an Argonaute protein; a cell line that overexpresses an Argonaute protein  
 CC ; an assay for identifying nucleic acid sequences for conferring a  
 CC particular phenotype in a cell; a nucleic acid composition composed of a  
 CC nucleic acid comprising an RNAi construct and a nucleic acid encoding an  
 CC Argonaute protein, where the RNAi construct comprises a nucleic sequence  
 CC encoding a single-strand short interfering RNA (siRNA); a pharmaceutical  
 CC composition comprising the nucleic acid composition; and a cell  
 CC expressing the nucleic acid composition. The methods and compositions of  
 CC the invention are useful for enhancing the effectiveness of an RNAi  
 CC therapeutic. The present sequence is that of a Caenorhabditis briggsae  
 CC Argonaute protein PIWI domain polypeptide, which was included in an  
 CC alignment of Ago protein polypeptides AEF73774-AEF73788 to examine  
 CC conservation of active site residues.

XX Sequence 94 AA;

Query Match 0.9%; Score 9; DB 10; Length 94;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRQGV 804

Db 70 IVVYRQGV 78

RESULT 11

AEF73783

ID AEF73783 standard; protein; 94 AA.

XX AEF73783;

XX 06-APR-2006 (first entry)

XX Caenorhabditis elegans ALG2 Argonaute protein PIWI domain polypeptide.

XX Argonaute; RNA interference; RNAi; drug screening; X-ray crystallography;

XX gene silencing.

XX

Caenorhabditis briggsae.

WO2006015258-A2.

09-FEB-2006.

28-JUL-2005; 2005WO-US027084.

28-JUL-2004; 2004US-0592227P.

29-JUL-2004; 2004US-0592269P.

(COLD-) COLD SPRING HARBOR LAB.

Joshua-Tor L, Song J, Hannon GU, Liu J, Carmell MA, Rivas F;  
Marsden C;

WPI; 2006-155768/16.

Crystalline Argonaute useful for identifying agent that binds Argonaute  
 protein, utilized for treating pancreatic cancer, treating hepatitis C  
 infection, cancer or inflammatory diseases.

Example 10; Fig 21; 215pp; English.

The present invention relates to Argonaute proteins, which are involved  
 in RNA interference (RNAi). The invention provides a crystalline  
 Argonaute protein, which comprises (i) an N-terminus, middle and PIWI  
 domain which form a crescent-shaped base, and (ii) a PAZ domain, which is  
 positioned above the crescent-shaped base, resulting in a cleft between  
 the crescent-shaped base and the PAZ domain. The structure of the full-  
 length Argonaute protein AEF73751 from the archaeobacterium Pyrococcus  
 furiosus (PfAgol), as determined by x-ray crystallography to 2.25 Angstrom  
 resolution, is provided. The invention also provides an isolated complex  
 comprising an Argonaute protein and a single-stranded RNA hybridized to  
 its target nucleic acid. In certain embodiments, the single-stranded RNA  
 is bound to the PAZ domain of the Argonaute protein, and may further  
 interact with the crescent-shaped base of the Argonaute protein. Also  
 claimed are: a method of determining the three-dimensional structure of  
 an Argonaute protein or its mutant, derivative, variant, analog, homolog,  
 sub-domain or fragment by alignment with the PfAgol amino acid sequence to  
 match homologous regions; a method of identifying an agent that binds an  
 Argonaute protein by applying a three-dimensional molecular modeling  
 algorithm to the atomic coordinates of an Argonaute protein to determine  
 the spatial coordinates of the binding pocket of the Argonaute protein,  
 and electronically screening the stored spatial coordinates of a set of  
 candidate agents against the spatial coordinates of the Argonaute protein  
 binding pocket; a computer-based method for the analysis of the  
 interaction of a molecular structure with an Argonaute protein; a data  
 array comprising the atomic coordinates of an Argonaute protein; a  
 computer-readable storage medium encoded with the atomic coordinates of  
 an Argonaute protein; an electronic representation of a domain or binding  
 site of the Argonaute protein; a method for obtaining a crystal of an  
 Argonaute protein; a method for identifying an agent that modulates the  
 activity of an RNAi construct by identifying an agent that modulates the  
 expression and/or activity of an Argonaute protein; a method for  
 identifying an agent that potentiates the activity of an RNAi construct  
 by identifying an agent that increases the expression and/or activity of  
 an Argonaute protein; a method of identifying an agent that modulates the  
 activity of an RNAi construct by examining a change in Argonaute protein  
 activity in the presence of a candidate agent; a composition for targeted  
 gene inhibition comprising an agent that modulates the RNase activity of  
 an Argonaute protein; a cell line that overexpresses an Argonaute protein  
 ; an assay for identifying nucleic acid sequences for conferring a  
 particular phenotype in a cell; a nucleic acid composition composed of a  
 nucleic acid comprising an RNAi construct and a nucleic acid encoding an  
 Argonaute protein, where the RNAi construct comprises a nucleic sequence  
 encoding a single-strand short interfering RNA (siRNA); a pharmaceutical  
 composition comprising the nucleic acid composition; and a cell  
 expressing the nucleic acid composition. The methods and compositions of  
 the invention are useful for enhancing the effectiveness of an RNAi  
 therapeutic. The present sequence is that of a Caenorhabditis briggsae  
 Argonaute protein PIWI domain polypeptide, which was included in an  
 alignment of Ago protein polypeptides AEF73774-AEF73788 to examine  
 conservation of active site residues.

CC conservation of active site residues.

XX Sequence 94 AA;

SQ

Query Match 0.9%; Score 9; DB 10; Length 94;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVS 804

DB 70 IVVYRDGVS 78

|||||

RESULT 12

ID AEF73782 standard; protein; 94 AA.

XX AEF73782;

AC

DT 06-APR-2006 (first entry)

XX

DE Caenorhabditis elegans ALG1 Argonaute protein PIWI domain polypeptide.

XX

KW Argonaute; RNA interference; RNAi; drug screening; X-ray crystallography;

KM gene silencing.

XX

OS Caenorhabditis briggsae.

XX

PN WO2006015259-A2.

XX

PD 09-FEB-2006.

XX

PP 28-JUL-2005; 2005WO-US027084.

XX

PR 28-JUL-2004; 2004US-0592297P.

PR 29-JUL-2004; 2004US-0592269P.

XX

PA (COLD-) COLD SPRING HARBOR LAB.

XX

PI Joshua-Tor L, Song J, Hannon GJ, Liu J, Carmell MA, Rivas F;

PI Marsden C;

XX

DR WPI; 2006-155768/16.

XX

PT Crystalline Argonaute useful for identifying agent that binds Argonaute

PT protein, utilized for treating pancreatic cancer, treating hepatitis C

PT infection, cancer or inflammatory diseases.

XX

PS Example 10; Fig 21; 215pp; English.

XX

CC The present invention relates to Argonaute proteins, which are involved

CC in RNA interference (RNAi). The invention provides a crystalline

CC Argonaute protein, which comprises (i) an N-terminus, middle and PIWI

CC domain which form a crescent-shaped base, and (ii) a PAZ domain, which is

CC positioned above the crescent-shaped base, resulting in a cleft between

CC the crescent-shaped base and the PAZ domain. The structure of the full-

CC length Argonaute protein AEF73751 from the archaeobacterium Pyrococcus

CC furiosus (PfAgol), as determined by x-ray crystallography to 2.25 Angstrom

CC resolution, is provided. The invention also provides an isolated complex

CC comprising an Argonaute protein and a single-stranded RNA hybridized to

CC its target nucleic acid. In certain embodiments, the single-stranded RNA

CC is bound to the PAZ domain of the Argonaute protein, and may further

CC interact with the crescent-shaped base of the Argonaute protein. Also

CC claimed are: a method of determining the three-dimensional structure of

CC an Argonaute protein or its mutant, derivative, variant, analog, homolog,

CC sub-domain or fragment by alignment with the PfAgol amino acid sequence to

CC match homologous regions; a method of identifying an agent that binds an

CC Argonaute protein by applying a three-dimensional molecular modeling

CC algorithm to the atomic coordinates of an Argonaute protein to determine

CC the spatial coordinates of the binding pocket of the Argonaute protein,

CC and electronically screening the stored spatial coordinates of a set of

CC candidate agents against the spatial coordinates of the Argonaute protein

CC binding pocket; a computer-based method for the analysis of the

CC interaction of a molecular structure with an Argonaute protein; a data

CC array comprising the atomic coordinates of an Argonaute protein; a

CC computer-readable storage medium encoded with the atomic coordinates of

CC an Argonaute protein; an electronic representation of a domain or binding

CC site of the Argonaute protein; a method for obtaining a crystal of an

CC Argonaute protein; a method for identifying an agent that modulates the

CC activity of an RNAi construct by identifying an agent that modulates the

CC expression and/or activity of an Argonaute protein; a method for

CC identifying an agent that potentiates the activity of an RNAi construct

CC by identifying an agent that increases the expression and/or activity of

CC an Argonaute protein; a method of identifying an agent that modulates the

CC activity of an RNAi construct by examining a change in Argonaute protein

CC activity in the presence of a candidate agent; a composition for targeted

CC gene inhibition comprising an agent that modulates the RNase activity of

CC an Argonaute protein; a cell line that overexpresses an Argonaute protein

CC ; an assay for identifying nucleic acid sequences for conferring a

CC particular phenotype in a cell; a nucleic acid composition composed of a

CC nucleic acid comprising an RNAi construct and a nucleic acid encoding an

CC Argonaute protein, where the RNAi construct comprises a nucleic sequence

CC encoding a single-strand short interfering RNA (siRNA); a pharmaceutical

CC composition comprising the nucleic acid composition; and a cell

CC expressing the nucleic acid composition. The methods and compositions of

CC the invention are useful for enhancing the effectiveness of an RNAi

CC therapeutic. The present sequence is that of a Caenorhabditis elegans

CC ALG1 Argonaute protein PIWI domain polypeptide, which was included in an

CC alignment of Ago protein polypeptides AEF73774-AEF73788 to examine

CC conservation of active site residues.

XX

SQ Sequence 94 AA;

Query Match 0.9%; Score 9; DB 10; Length 94;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 IVVYRDGVS 804

DB 70 IVVYRDGVS 78

|||||

RESULT 13

ID ADB17506 standard; protein; 389 AA.

XX

AC ADB17506;

XX

DT 20-NOV-2003 (first entry)

XX

DE Wheat post-transcriptional gene silencing Argonaute homolog protein #5.

XX

KW Wheat; post-transcriptional gene silencing; PTGS; RNA degradation;

KW transgene expression; plant; transgenic plant; transposon activity;

KW meristem activity; architecture; development; proliferation;

XX callus propagation.

OS Triticum aestivum.

XX

PN US2003077623-A1.

XX

PD 24-APR-2003.

XX

PF 17-JUN-2002; 2002US-00174363.

XX

PR 18-JUN-2001; 2001US-0298973P.

XX

PA (BUTL/) BUTLER K H.

PA (CAHO/) CAHOON R E.

PA (HARV/) HARVELL L T.

PA (RAFA/) RAFALSKI J A.

PA (SAKA/) SAKAI H.

XX

PI Butler KH, Cahoon RE, Harvell LT, Rafalski JA, Sakai H;

DR WPI; 2003-644827/61.

DR N-PSDB; ADB17505.  
 XX  
 PT New polynucleotides encoding polypeptides comprising post-transcriptional  
 PT gene silencing activity useful in creating transgenic plants having  
 PT enhanced or deficient post-transcriptional gene silencing.  
 XX  
 PS Disclosure; Page 85-86; 107pp; English.  
 XX  
 CC The invention discloses an isolated polynucleotide encoding a polypeptide  
 CC having post-transcriptional gene silencing (PTGS) activity. PTGS operates  
 CC at the level of sequence specific RNA degradation and down regulates  
 CC transgene expression in plants. Also claimed is a recombinant DNA  
 CC construct comprising the polynucleotide operably linked to at least one  
 CC regulatory sequence, transforming a cell with the polynucleotide, a cell  
 CC comprising the recombinant construct, producing a polypeptide having post  
 CC -transcriptional gene silencing activity, producing a plant, a plant and  
 CC polypeptide comprising post-transcriptional DNA construct and an isolated  
 CC disclosed are methods for obtaining a nucleic acid fragment encoding all,  
 CC or a substantial portion of, the amino acid sequence encoding a  
 CC polypeptide involved in post-transcriptional gene silencing. Positive  
 CC selection of a transformed cell and altering the level of expression of a  
 CC polypeptide involved in post-transcriptional gene silencing in a host  
 CC cell. The polynucleotides encoding polypeptides involved in gene  
 CC silencing are useful in altering the post-transcriptional gene silencing  
 CC pathway in plant cells. These are specifically useful in creating  
 CC transgenic plants where the levels of polypeptides involved in post-  
 CC transcriptional gene silencing are altered with respect to non-transgenic  
 CC plants which would result in plants with an enhanced or deficient post-  
 CC transcriptional gene silencing. The polynucleotides may also be used to  
 CC manipulate transposon activity, meristem activity, plant architecture and  
 CC callus development, proliferation of undifferentiated plant cells in culture for  
 CC genes that they are part of and as markers for traits linked to those  
 CC genes. The sequence presented is a polypeptide having post-  
 CC transcriptional gene silencing activity.  
 XX  
 SQ Sequence 389 AA;  
 Query Match 0.9%; Score 9; DB 6; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 677 LKINAKLGG 685  
 Db |||||  
 89 LKINAKLGG 97  
 RESULT 14  
 AEC75730  
 ID AEC75730 standard; protein; 389 AA.  
 XX  
 AC AEC75730;  
 XX  
 DT 17-NOV-2005 (first entry)  
 XX  
 DE Wheat Argonaute (AGO1) homolog FIS clone protein, SEQ ID NO: 50.  
 XX  
 KW Gene silencing; transgenic plant; cell proliferation; agriculture;  
 KW crop improvement; Argonaute protein.  
 XX  
 OS Triticum aestivum.  
 XX  
 FN US2005204427-A1.  
 XX  
 PD 15-SEP-2005.  
 XX  
 PF 30-MAR-2005; 2005US-00093888.  
 XX  
 PR 18-JUN-2001; 2001US-0298973P.  
 PR 17-JUN-2002; 2002US-00174363.  
 XX  
 PA (BUTL/) BUTLER K H.

PA (CAHO/) CAHOON R E.  
 PA (HARV/) HARVELL L T.  
 PA (RAFA/) RAFALSKI J A.  
 PA (SAKA/) SAKAI H.  
 XX  
 PI Butler KH, Cahoon RE, Harvell LT, Rafalski JA, Sakai H;  
 XX  
 DR WPI; 2005-618219/63.  
 DR N-PSDB; AEC75729.  
 XX  
 PT New polynucleotide encoding a polypeptide having post-transcriptional  
 PT gene silencing activity, useful for manipulating plant architecture and  
 PT development, or for propagating callus.  
 XX  
 PS Disclosure; SEQ ID NO 50; 107pp; English.  
 XX  
 CC The invention relates to a polynucleotide encoding a polypeptide having  
 CC post-transcriptional gene silencing (PTGS) activity. The invention also  
 CC relates to a method for producing transgenic plants. The sequences and  
 CC method of the invention are useful for manipulating transposon activity,  
 CC meristem activity, plant architecture and development or proliferation of  
 CC undifferentiated plant cells in culture and for propagating callus. The  
 CC present sequence is the wheat Argonaute (AGO1) homolog protein. This  
 CC sequence is involved in PTGS activity.  
 XX  
 SQ Sequence 389 AA;  
 Query Match 0.9%; Score 9; DB 9; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 677 LKINAKLGG 685  
 Db |||||  
 89 LKINAKLGG 97  
 RESULT 15  
 ADQ80691  
 ID ADQ80691 standard; protein; 850 AA.  
 XX  
 AC ADQ80691;  
 XX  
 DT 21-OCT-2004 (first entry)  
 XX  
 DE Arabidopsis thaliana TFL1-binding protein #3.  
 XX  
 KW TFL1-binding protein; plant growth control; biotechnology;  
 KW fishing industry; screening.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 783..792  
 FT /note= "Encoded by nucleotides 2347..2445"  
 XX  
 FN JP2004208572-A.  
 XX  
 PD 29-JUL-2004.  
 XX  
 PF 27-DEC-2002; 2002JP-00381220.  
 XX  
 PR 27-DEC-2002; 2002JP-00381220.  
 XX  
 PA (DOKU-) DOKURITSU GYOSEI HOJIN KAGAKU GIJUTSU SH.  
 XX  
 DR WPI; 2004-538191/52.  
 DR N-PSDB; ADQ80690.  
 XX  
 PT Novel AGO protein, comprising PA2 and Piwi domain, specifically binding  
 PT to TFL1 protein that is derived from Arabidopsis thaliana, useful for  
 PT controlling growth phase of plant.  
 XX  
 PS Claim 2; SEQ ID NO 6; 104pp; Japanese.



XX The invention comprises the amino acid and coding sequences of  
CC Arabidopsis thaliana proteins which specifically bind to TFL1 protein.  
CC The amino acid and coding sequences of the invention are useful for  
CC screening TFL1 family proteins. The DNA and protein sequences of the  
CC invention are also useful for controlling the growth phase of a plant,  
CC and as a reagent for research in biotechnology and fishing industry. The  
CC present amino acid sequence represents an Arabidopsis thaliana TFL1-  
CC binding protein of the invention.  
XX  
SQ Sequence 850 AA;  
  
Query Match 0.9%; Score 9; DB 8; Length 850;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 677 LKINAKLGG 685  
|||  
Db 569 LKINAKLGG 577  
  
Search completed: July 5, 2006, 13:53:28  
Job time : 199 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 5, 2006, 22:39:47 ; Search time 9528 Seconds  
(without alignments)  
8979.490 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 5349  
Sequence: 1 MSSNPFLEKGFVHSLDPE.....RHEMFLOTNVYKVGMSFA 1020

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/abes/ABSSWEB\_epool/US10645746/runat\_05072006\_110146\_5084/app\_query.fasta\_1  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes05p  
-USER=US10645746 @CGN 1\_1\_9528 @runat\_05072006\_110146\_5084 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est3: \*  
3: gb\_est4: \*  
4: gb\_est5: \*  
5: gb\_est6: \*  
6: gb\_est7: \*  
7: gb\_est8: \*  
8: gb\_est9: \*  
9: gb\_est10: \*  
10: gb\_est11: \*  
11: gb\_est12: \*  
12: gb\_est13: \*  
13: gb\_est14: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1189	22.2	796	2	BJ155598
C 2	1173	21.9	786	2	BJ143897
C 3	1088	20.3	633	2	BJ122122
C 4	1063	19.9	700	2	BJ150328

C 5	1045	19.5	710	2	BJ132853
C 6	1034	19.3	700	2	BJ155451
C 7	1033	19.3	635	2	BJ755060
C 8	1009	18.9	677	2	BJ779221
C 9	999	18.7	565	4	CB401772
C 10	995	18.6	583	2	BJ116268
C 11	983	18.4	554	4	CB398209
C 12	944	17.6	557	2	BJ105221
C 13	928	17.3	534	2	BJ127233
C 14	911	17.0	539	2	BJ127087
C 15	670	12.5	378	4	C62850
C 16	617	11.5	360	4	C60407
C 17	611	11.4	360	9	D35146
C 18	606	11.3	356	10	Z14900
C 19	605	11.3	360	4	C65335
C 20	605	11.3	360	9	D35838
C 21	590	11.0	373	4	C60787
C 22	584	10.9	370	1	AV191613
C 23	576	10.8	376	4	C60885
C 24	574	10.7	449	9	D27223
C 25	574	10.7	2465	14	AY417690
C 26	574	10.7	2465	14	AY417692
C 27	554	10.4	3778	6	AK035042
C 28	553	10.3	3845	6	AK030018
C 29	550.5	10.3	3160	6	AK154436
C 30	550.5	10.3	3160	6	AK171118
C 31	537.5	10.0	3127	6	AK083652
C 32	535	10.0	3606	6	AK080954
C 33	532	9.9	340	9	D27220
C 34	529.5	9.9	3123	13	CL969749
C 35	528	9.9	3509	6	BC073067
C 36	526	9.8	334	10	M89235
C 37	526	9.8	421	9	D27221
C 38	517.5	9.7	3188	6	CNS0AAU5
C 39	510	9.5	2955	13	CL965306
C 40	496.5	9.3	3168	13	CL963072
C 41	495	9.3	3222	13	CL969750
C 42	486	9.1	3336	13	CL967201
C 43	479	9.0	375	9	D32492
C 44	479	9.0	2522	14	DQ040105
C 45	476	8.9	295	9	D27222

#### ALIGNMENTS

RESULT 1	BJ155598/c	796 bp	mrna	linear	EST 24-JAN-2002
LOCUS	BJ155598	unpublished oligo-capped cDNA library, C. elegans LI stage			
DEFINITION	Caenorhabditis elegans cDNA clone YK1344B04 3', mRNA sequence.				
ACCESSION	BJ155598				
VERSION	BJ155598.1	GI:18323583			
KEYWORDS	EST.				
SOURCE	Caenorhabditis elegans				
ORGANISM	Caenorhabditis elegans				
REFERENCE	1 (bases 1 to 796)				
AUTHORS	Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. and Sugano, S.				
TITLE	A complementary view of the C.elegans genome				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.				
FEATURES	Location/Qualifiers				
source	1..796 /organism="Caenorhabditis elegans" /mol_type="mRNA"				

```
/strain="N2"
/db_xref="taxon:6239"
/clone="yk134504"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"

ORIGIN
Alignment Scores:
Pred. No.: 7,08e-125 Length: 796
Score: 1189.00 Matches: 225
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 99.6% Mismatches: 0
Query Match: 22.2% Indels: 0
DB: 2 Gaps: 0

US-10-645-746-3 (1-1020) x BJ155598 (1-796)

QY 795 HisIleValValTyRArgAspGlyValSerAspSerGluMetLeuArgValSerHisAsp 814
DB 795 CATATTGTAGTCTATCGACACGGAGTNACGATTCGGAGATGCTACGTGTTAGTCATGAT 736

QY 815 GluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAsp 834
DB 735 GAGCTTCGATCTTTAAAGCCGAAGTAAACAATTCATGTCGAAACGGAGTGGAGAGAT 676

QY 835 ProGluProLysTyRThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArg 854
DB 675 CCAGAGCCGAGTACACGTTTCATTGTGATTCAGAAAGACACATACACGATTGCTTCCA 616

QY 855 ArgMetGluLysAspLysProValValLeuLysAspLeuThrProAlaGluThrAspVal 874
DB 615 AGAATGGAAAAAGATAAGCCAGTCGTGTCATTAAGATCTTACTCTCTGCTGAAACAGATGTC 556

QY 875 AlaValAlaAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyIle 894
DB 555 GCTGTGCTGCTGTTAAACATGGGAGGAGGATATGAAGAAGACAAAGAAACTGGAAAT 496

QY 895 ValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyRlysPheAsp 914
DB 495 GTGAACCCATCATCCGAAACAACATGTCGATAAATTCATGTTTCGAAATACAAATTCGAT 436

QY 915 PhePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyRThrVal 934
DB 435 TTTTCTCTGTCATCTCATGTCGTCCTGTGTGTATCATCTCGCCAGGACATTACACTGTT 376

QY 935 MetTyRAspLysGlyMetSerGlnAspGluValTyRLeuMetThrTyRglyLeuAla 954
DB 375 ATGTATGACGATAAAGGAATGAGCCAGATGAAGTCTATAAATGACCTACCGACTTGCT 316

QY 955 PheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyRAlaHis 974
DB 315 TTTCTCTCTGCTAGATGTCGAAAAACCATCTGTCCTGTCGTCGTCGTCAT 256

QY 975 LeuSerCysGluLysAlaLysGluLeuTyRArgThrTyRlyLeuGluHisTyRileGlyAsp 994
DB 255 TTATCATGTGAAAAACCGAAGAGCTTTTATCGAATCTTACAGGAACATTACATCGGTGAC 196

QY 995 TyRAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnThrAsnValLysTyR 1014
DB 195 TATGCACAGCCAGGACTCGACACGAAATGGAAACATTTTCTCCAAACTAACTGGAAGTAC 136

QY 1015 ProGlyMetSerPheAla 1020
DB 135 CCTGGAATGCTGTCGCA 118
```

```
RESULT 2
BJ143897/c
LOCUS BJ143897 unpublished oligo-capped cDNA library, C. elegans L1 stage
DEFINITION Caenorhabditis elegans cDNA clone yk1201c08 3', mRNA sequence.
```

```
ACCESSION BJ143897
VERSION BJ143897.1 GI:18304063
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 786)
Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
source
1..786
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1201c08"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"
```

## ORIGIN

```
Alignment Scores:
Pred. No.: 4,75e-123 Length: 786
Score: 1173.00 Matches: 222
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 21.9% Indels: 0
DB: 2 Gaps: 0

US-10-645-746-3 (1-1020) x BJ143897 (1-786)

QY 799 TyRArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSer 818
DB 785 TATCGAGACGGAGTNAGCGATTCGGAGATGCTACGTGTTAGTCATGATGAGTTCGATCT 726

QY 819 LeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLys 838
DB 725 TTAATAACGCGAAGTAAACAATTCATGTCGGAACGGGATGGAGAAGATCCAGAGCCGAAG 666

QY 839 TyRThrPheIleValIleGlnLysArgHisAsnThrArgLeuArgMetGluLys 858
DB 665 TACACGTTTCATTGTGATTCAGAAAAAGACACAAATACACGATTCCTTCGAAAGAATGGAATAA 606

QY 859 AspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAlaValAlaAla 878
DB 605 GATAAGCCAGTGGTCAATTAAGATCTTACTCTGCTGAAACAGATGTCGCTGTCGTGCT 546

QY 879 ValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyIleValAsnProSer 898
DB 545 GTTAACAATCGGAGGAGGATATGAAGAAGCAAAAGAAACTGGAATTTGTGAACCCATCA 486

QY 899 SerGlyThrThrValAspLysLeuIleValSerLysTyRlysPheAspPheLeuAla 918
DB 485 TCCGGAAACAACCTGTGGATAAACTTATCGTTTCGAAATACAAATTCGATTTTCTTGGCA 426

QY 919 SerHisGlyValLeuGlyThrSerArgProGlyHisTyRThrValMetTyRAspAsp 938
DB 425 TCTCATCATGTCGTCCTTGGTACATCTCTGTCAGGACATTCACCTGTTATGATGACGAT 366

QY 939 LysGlyMetSerGlnAspGluValTyRlysMetThrTyRglyLeuAlaPheLeuSerAla 958
```

```

Db      365  AAGGAATGAGCCAGAGTCTATATAAATGACCTACGACTTGCTTTCTCTGCT 306
Qy      959  ArgCysArgLysProIleSerLeuProValHisTyrAlaHisLeuSerCysGlu 978
Db      305  AGATGTGGAAGAACCATCTCGTTCGCTGTTCCGGTTCATTATGCTCATATCATGTGA 246
Qy      979  LysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnPro 998
Db      245  AAGCGAAGAGCTTTATCGAACTTACAGGACATTACATCGGTGACTATGCACGCCA 186
Qy      999  ArgThrArgHisGluMetGluHisPheLeuGlnThrAsnValLysTyrProGlyMetSer 1018
Db      185  CGGACTCGACAGAAATGGAACATTTCTTCCAACTACGCTGAAGTACCTCGGAATGCG 126
Qy      1019 PheAla 1020
Db      125  TTCGCA 120

RESULT 3
BJ122122
LOCUS
DEFINITION
  BUI22122 unpublished oligo-capped cDNA library, C. elegans L1 stage
  BUI22122
  BUI22122.1 GI:18282260
ACCESSION
  BUI22122
VERSION
  BUI22122.1
KEYWORDS
  EST.
SOURCE
  Caenorhabditis elegans
  Caenorhabditis elegans
  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
  Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
  1 (bases 1 to 633)
  Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
  and Sugano, S.
  A complementary view of the C.elegans genome
  Unpublished (2002)
  Contact: Tadasu Shin-i
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
  Location/Qualifiers
    1..633
    /organism="Caenorhabditis elegans"
    /mol_type="mRNA"
    /strain="N2"
    /db_xref="taxon:6239"
    /clone="yk1284h06"
    /sex="hermaphrodite"
    /tissue_type="whole animal"
    /dev_stage="L1"
    /clone_lib="unpublished oligo-capped cDNA library, C.
    elegans L1 stage"

FEATURES
  source
  1..633
  Alignment Scores:
  Pred. No.: 1.9e-113 Length: 633
  Score: 1088.00 Matches: 209
  Percent Similarity: 99.5% Conservative: 1
  Best Local Similarity: 99.1% Mismatches: 1
  Query Match: 20.3% Indels: 0
  DB: 2 Gaps: 0

ORIGIN
  US-10-645-746-3 (1-1020) x BJ122122 (1-633)

Qy      453  LeuLysMetLeuLysGluLeuAspPheSerSerGluLeuAsnPheValGluArgPhe 472
Db      1  CTAAAAATGCTGAAGAAATGGATTTCTCTTCGAAGAGCTAAATTTGTGGAAGATTT 60
Qy      473  GlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLysGluProMet 492
Db      61  GGATATGCTCCAACTTCAGATCGAATGTCAGGAAGAGGTTTGAAGAGCCATG 120

```

```

Qy      493  LeuValAsnSerValAsnGluGlnIleLysMetThrProValIleArgGlyPheGlnGlu 512
Db      121  CTTGTGAATAGTGTAAATGAACAAATTAATAATGACACAGTGATTCGTGGATTTCAAGAA 180
Qy      513  LysGlnLeuAsnValProGluLysGluLeuCysCysAlaValPheValValAsnGlu 532
Db      181  AAACAATGGAATGTGGTTCCGAAAAAGAACTTTGCTGTGTCTGTTTTGTAGCCCAACGAA 240
Qy      533  ThrAlaGlyAsnProCysLeuGluGluAsnAspValValLysPheTyrThrGluLeuIle 552
Db      241  ACAGCGGGAATCCATGCTTAGAAGAGACGACGTTGTTAAGTTCTACACCGAACTAATT 300
Qy      553  GlyGlyCysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyValGlnSer 572
Db      301  GGTGGTTGCAAGTTCCGTGGAATACGAATGGTGCCAATGAAACACAGAGGAGCGCAATCT 360
Qy      573  IleMetTyrAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThr 592
Db      361  ATTATGTACACGACGACGAGAAAAATGAATATGCCCTTCTACAAAAATTGTACATAATACC 420
Qy      593  GlyIleGlyArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuPro 612
Db      421  GGAATCGGTAGATTTGAATAGCCGCAACAGAGCGGAAGATATGTTTGACGCTCTTCCC 480
Qy      613  AspLysGluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyr 632
Db      481  GATAAAGAACACCAAGTCTTAATGTTTCATTATCATTTCCAAACGACAACTGAATGCTTAC 540
Qy      633  GlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSer 652
Db      541  GGTGTTGTGAAACATTATTTCGATCACCATCGGTGTAGCTAATCATCAGCATATTTACTTCT 600
Qy      653  GluThrValThrLysAlaLeuAlaSerLeuArg 663
Db      601  GAAACAGTCACAAAGCTTTGGCATCTACTAAGG 633

```

BUI50328 700 bp mRNA linear EST 24-JAN-2002  
 BUI50328 unpublished oligo-capped cDNA library, C. elegans L1 stage  
 Caenorhabditis elegans cDNA clone yk1284h06 3', mRNA sequence.

BUI50328  
 BUI50328.1 GI:18318313  
 EST.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Caenorhabditis elegans  
 Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE  
 1 (bases 1 to 700)  
 Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.  
 and Sugano, S.

A complementary view of the C.elegans genome  
 Unpublished (2002)

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES  
 1..700  
 Location/Qualifiers

/organism="Caenorhabditis elegans"

/mol\_type="mRNA"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk1284h06"

/sex="hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="L1"

/clone\_lib="unpublished oligo-capped cDNA library, C.  
 elegans L1 stage"

## ORIGIN

Alignment Scores:  
Pred. No.: 1,64e-110 Length: 700  
Score: 1063.00 Matches: 200  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 19.9% Indels: 0  
DB: 2 Gaps: 0

US-10-645-746-3 (1-1020) x BJ150328 (1-700)

```
QY 821 SerGluValLysGluPheMetSerGluArgAspGlyGluAspProGluProLysTyrThr 840
DB 699 AGCGAAGTAAACAATTCATGTCGGAACGGGATGGAGAAGATCCAGAGCGGAAGTACAG 640
QY 841 PheIleValIleGlnLysArgHisAenThrArgLeuLeuArgMetGluLysAspLys 860
DB 639 TTCATGTGATTCAGAAAGACACAAATACACCATTCGTTCCGAAGATGGAAAAAGATAAG 580
QY 861 ProValValAsnLysAspLeuThrProAlaGluThrAspValAlaValAlaValLys 880
DB 579 CCAGTGGTCAATAAAGATCTTACTCTGCTGGAACAGATGTCGCTGTTGCTGCTGTTAAA 520
QY 881 GlnTTPGluGluAspMetLysGluSerLysGluThrGlyLleValAsnProSerSerGly 900
DB 519 CAATGGGAGGAGATATGAAGAAGAAAGCAAGAAACTGGGAATTTGTGAACCATCATCCGGA 460
QY 901 ThrThrValAspLysLeuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHis 920
DB 459 ACAACTGTGGATAAATTCATGTTCCGAATAACAATTCGATTTTCTTGGCGTCTCAT 400
QY 921 HisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGly 940
DB 399 CATGGTGTCTTGGTACATCTCTGCTCAGGACATTAACCTGTTATGTATGACGATAAAGGA 340
QY 941 MetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCys 960
DB 339 ATGACCAAGATGAAGTCTATAAATGACCTACGGACTTGTCTTCTCTGCTAGATGT 280
QY 961 ArgLysProIleSerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAla 980
DB 279 CGAAAAACCATCTCGTTGCTCTGCTGCTGCTATTCATTCATTCATGCAAAAAAGCG 220
QY 981 LysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThr 1000
DB 219 AAAGAGCTTTATCGAACTTTACAAGGAACATTAACATCGGTGACTATGCACAGCCGAGCT 160
QY 1001 ArgHisGluMetGluHisPheLeuGlnThrAenValLysTyrProGlyMetSerPheAla 1020
DB 159 CGACACGAATGGAACATTTTCTCCAACTAACTGTAAGTACCTGGGAATGTCGTTCCGA 100
```

RESULT 5  
BJ132853/c 710 bp mRNA linear EST 23-JAN-2002  
LOCUS  
DEFINITION  
Caenorhabditis elegans cdna clone yk1070f12 3', mRNA sequence.

ACCESSION  
BJ132853  
VERSION  
BJ132853.1 GI:18293010  
KEYWORDS  
EST.  
SOURCE  
Caenorhabditis elegans  
Caenorhabditis elegans  
ORGANISM

REFERENCE  
1 (bases 1 to 710)  
Kohara.Y., Shin-I.T., Thierry-Mieg.J., Thierry-Mieg.D., Suzuki.Y.  
and Sugano.S.

TITLE  
A complementary view of the C.elegans genome  
JOURNAL  
Unpublished (2002)  
COMMENT  
Contact: Tadao Shin-I  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

FEATURES  
source

Location/Qualifiers  
1..710  
/organism="Caenorhabditis elegans"  
/mol\_type="mRNA"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk1070f12"  
/sex="hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="L1"  
/clone\_lib="unpublished oligo-capped cdna library, C.  
elegans L1 stage"

## ORIGIN

Alignment Scores:  
Pred. No.: 1,95e-108 Length: 710  
Score: 1045.00 Matches: 196  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 19.5% Indels: 0  
DB: 2 Gaps: 0

US-10-645-746-3 (1-1020) x BJ132853 (1-710)

```
QY 825 GlnPheMetSerGluArgAspGlyGluAspProGluProLysTyrThrPheIleValIle 844
DB 705 CAATTCATGTCGGAACGGGATGGAGAAGATCCAGAGCCGGAAGTACACGTTTCATTTGAT 646
QY 845 GlnLysArgHisAenThrArgLeuLeuArgMetGluLysAspLysProValValAen 864
DB 645 CAGAAAGACACAATACACGATTCGTTCCGAAGATGGAAAAAGATAAGCCAGTGGTCAAT 586
QY 865 LysAspLeuThrProAlaGluThrAspValAlaValAlaValLysGlnTyrGluGlu 884
DB 585 AAAGATCTTACTCTGCTGGAACAGATGTCGCTGTTGCTGCTGCTCACAACATGGGAGGAG 526
QY 885 AspMetLysGluSerLysGluThrGlyLleValAenProSerSerGlyThrThrValAap 904
DB 525 GATATGAAGAAAGCAAGAACTGGAATTTGTAACCCCATCATCCGGAACAACATGTGGAT 466
QY 905 LysIleuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHisGlyValLeu 924
DB 465 AAACATTCGTTTCGAAATACAAATTCGATTTTCTTGGCATCTCATCATGTTGCTCT 406
QY 925 GlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAap 944
DB 405 GGTACATCTCGTCCAGGACATTTACACTGTTATGATATGACGATAAAGGAATGAGCCAAGAT 346
QY 945 GluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIle 964
DB 345 GAAGTCTATAAATGACCTACGAGTCTGCTTTTCTCTCTGCTAGATGTGCAAAACCCATC 286
QY 965 SerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyr 984
DB 285 TCGTTGCTGTTCTCCGTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 226
QY 985 ArgThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThrArgHisGluMet 1004
DB 225 CGAAGTCTACAGGAACATTTACATCGTGTGACTATGCACAGCCAGGACTCGACACGAAATG 166
QY 1005 GluHisPheLeuGlnThrAenValLysTyrProGlyMetSerPheAla 1020
DB 165 GAACATTTTCTCCAAACTAACGTGAAGTACCTCGGAATGTGCTTCGCA 118
```

## RESULT 6

BJ155451/c 700 bp mRNA linear EST 24-JAN-2002  
LOCUS  
DEFINITION  
BJ155451 unpublished oligo-capped cdna library, C. elegans L1 stage  
Caenorhabditis elegans cdna clone yk1342d11 3', mRNA sequence.  
ACCESSION  
BJ155451

```

VERSION BJ155451.1 GI:18323436
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE 1 (bases 1 to 700)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
    source
        1..700
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk1342d11"
            /sex="hermaphrodite"
            /tissue_type="whole animal"
            /dev_stage="L1"
            /clone_lib="unpublished oligo-capped cDNA library, C.
            elegans L1 stage"
ORIGIN
Alignment Scores:
Pred. No.: 3,49e-107 Length: 700
Score: 1034.00 Matches: 194
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 19.3% Indels: 0
DB: 2 Gaps: 0
US-10-645-746-3 (1-1020) x BJ155451 (1-700)
QY 827 MetSerGluArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGlnLys 846
DB 699 ATGTCGGAACGGATGGAGAAGATCCAGACCCAGTACAGTTCATGTGATTCAGAA 640
QY 847 ArgHisAsnThrArgLeuLeuArgMetGluLysAspLysProValValAsnLysAsp 866
DB 639 AGACACATACACGATGCTTCGAGAGATGGAAAAGATNAGCCAGTGTCAATAAGAT 580
QY 867 LeuThrProAlaGluThrAspValAlaValAlaAlaValLysGlnTrpGluGluAspMet 886
DB 579 CTTACTCTCTGCTGAAACAGATGTCGCTGTGCTGTTAAACAATGGGAGGAGATATG 520
QY 887 LysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeu 906
DB 519 AAAGNAGCAAGCAAGAAATGGAAATGTGAACCCCATCATCCGGAACAATGTGGATAAAT 460
QY 907 IleValSerLysTyrLysPheAspPhePheLeuAlaSerHisGlyValLeuGlyThr 926
DB 459 ATGCTTCGGAATACAAATTCGATTTTCTTGCGCATCTCATCATGTGTCTTGTGACA 400
QY 927 SerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluVal 946
DB 399 TCTCGTCCAGGACATACACTGTATGTATGACGATTAAGGAATGAGCCAAAGATGAATC 340
QY 947 TyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeu 966
DB 339 TATAAATGACCTACCGACTGCTTTCTCTCTGCTAGATGTCGAAACCCATCTCGTGTG 280
QY 967 ProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThr 986
DB 279 CCTGTTCCGGTTCATATGCTCATTTATCATGTGAAAGGCGAAGAGCTTTATCGAACT 220

```

```

QY 987 TyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHis 1006
DB 219 TACAGGACATTTACATCGTGCATATGCACAGCCAGGACTGCACACGAAATCGAATCAT 160
QY 1007 PheLeuGlnThrAsnValLysTyrProGlyMetSerPheAla 1020
DB 159 TTCTCCAAACTAACTGAAGTACCCCTGGAAATGTCGTTCGCA 118
RESULT 7
LOCUS BJ755060
DEFINITION BJ755060 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone yk1409d07 5', mRNA sequence.
ACCESSION BJ755060
VERSION BJ755060.1 GI:47594822
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE 1 (bases 1 to 635)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
    source
        1..635
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk1409d07"
            /sex="Hermaphrodite"
            /tissue_type="whole animal"
            /dev_stage="varied"
            /clone_lib="unpublished oligo-capped cDNA library"
ORIGIN
Alignment Scores:
Pred. No.: 3,89e-107 Length: 635
Score: 1033.00 Matches: 194
Percent Similarity: 99.5% Conservative: 0
Best Local Similarity: 99.5% Mismatches: 1
Query Match: 19.3% Indels: 0
DB: 2 Gaps: 0
US-10-645-746-3 (1-1020) x BJ755060 (1-635)
QY 1 MetSerSerAsnPheProGluLeuLysGlyPheTyrArgHisSerLeuAspProGlu 20
DB 51 ATGTCCTCGAATTTTCCCGAATTTGGAAAAGGATTTATCGTCATCTCTCGATCCGAG 110
QY 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysVal 40
DB 111 ATGAAATGGCTTCGCGAGGCCCATCTGGTAAATGCGACGCAAAATTTCTATGAGAAAGTA 170
QY 41 LeuLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGlu 60
DB 171 CTTCTTTTGGTAAATTTGGTTCAAGTTCTCAGCAAAATTTACGATCGGAAATCTACGAG 230
QY 61 TyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLys 80
DB 281 TATGAAGTCAAAATGCAAAAGGAAGTATTGAATAGAAACCCAGCAAAACCTTTCCCAAAA 290
QY 81 LysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGlu 100

```

```

Db      291  AGACAGAAATTCATTCGATCGTGCAGAACTCTTCTGGCTACATCTTGGCATGAG 350
      |||
Qy      101  LysLysGlnThrAspPheLeuGluAspTyrValPheAspGluLysAspThrValTyr 120
      |||
Db      351  AAGAGCAGACAGATTTTATCTCGAAGACTATGTTTTTGTGATGAAAAGCAGACTGTTTAT 410
      |||
Qy      121  SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysValValLys 140
      |||
Db      411  AGTGTGTGCGACTGCACTGTGCATCAAAAATCTCGTTTCGGAGAAAGTAGTAAAA 470
      |||
Qy      141  LysAepSerGluLysLysAspGluLysAspLeuGluLysLysLeuLysLysLeuTyrThrMetile 160
      |||
Db      471  AAGGATTCGGAGAAAAGATGAAAGGATTTGGAGAAAAAATCTTATACAAATGATA 530
      |||
Qy      161  LeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGluLysAspGlu 180
      |||
Db      531  CTTACCTATCGTAAAAAATTTCACTGAACTTTAGTCGAGAAAAATCCGGAAAAAGACGAA 590
      |||
Qy      181  GluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGln 195
      |||
Db      591  GAAGCGAATCGGAGTTACAAATTCCTGAAGAATGTTATGACCCAG 635
      |||

```

```

RESULT 8
BJ779221/c
LOCUS      677 bp mRNA linear EST 25-MAY-2004
DEFINITION BJ779221 unpublished oligo-capped cDNA library Caenorhabditis
ACCESSION  BJ779221
VERSION     BJ779221.1 GI:47654855
KEYWORDS   EST.
SOURCE     Caenorhabditis elegans
ORGANISM   Caenorhabditis elegans
REFERENCE  1 (bases 1 to 677)
AUTHORS    Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE      A complementary view of the C.elegans genome
JOURNAL    Unpublished (2002)
COMMENT    Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tschini@genes.nig.ac.jp.
Location/Qualifiers
1. .677
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1409d07"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="unpublished oligo-capped cDNA library"

```

ORIGIN

```

Alignment Scores:
Pred. No.:      2,44e-104      Length:      677
Score:          1009.00      Matches:    189
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    18.9%      Indels:    0
DB:             2      Gaps:    0

US-10-645-746-3 (1-1020) x BJ779221 (1-677)

```

```

Qy      832  GlyGluAspProGluProLysTyrThrPheLeuValLysGlnLysArgHisAsnThrArg 851
      |||
Db      675  GGAGAAGATCCAGAGCCGGAAGTACACGTTTCATTGTGATTCAGAAAAGACACATACAGGA 616
      |||

```

```

Qy      852  LeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGlu 871
      |||
Db      615  TTGCTTCGAAGAATCGAAAAAGATAAGCCAGTGGTCAATAAAAGATCTTTACTCTGCTGAA 556
      |||
Qy      872  ThrAspValAlaValAlaValLysGlnTspGluGluAspMetLysGluSerLysGlu 891
      |||
Db      555  ACAGATGTGCGCTGTTGCTGCTTTAAACATCGGAGGAGGATATGAAAGAAAGCAAGAA 496
      |||
Qy      892  ThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyr 911
      |||
Db      495  ACTCGAAATTGTGAACCCATCATCCCGAACCACTGTGGATAAACTTATCGTTTCGAATAC 436
      |||
Qy      912  LysPheAspPhePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHis 931
      |||
Db      435  AAATTCGATTTTTTCTTGGCATCTCATCATGTGTGCTTGGTACATCTCGTCAGACAT 376
      |||
Qy      932  TyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyr 951
      |||
Db      375  TACACTGTTATGTATGACGATAAAGGAATGACCAAGATGAGTCTATATAAATGACCTAC 316
      |||
Qy      952  GlyLeuAlaPheLeuSerAlaArgCysArgLysProLysSerLeuProValProValHis 971
      |||
Db      315  GGACTTGGCTTTCTCTGTAGATGTGAAAAACCATCTCGTTGCTTCCGCTTCAT 256
      |||
Qy      972  TyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyr 991
      |||
Db      255  TATGCTCATTTATCATGTGAAAAAGCAAGAGCTTTATCGAACTTACAAAGAACATTAAC 196
      |||
Qy      992  IleGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnThrAsn 1011
      |||
Db      195  ATCGTGACTATGCACAGCCAGGACTCGACACGAATGGAACATTTCTCCAAACTAAC 136
      |||
Qy      1012  ValLysTyrProGlyMetSerPheAla 1020
      |||
Db      135  GTGAAGTACCCTCGGAATGCTGTTCCGA 109
      |||

```

RESULT 9

```

CB401772
LOCUS      565 bp mRNA linear EST 15-MAY-2003
DEFINITION CB401772_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
ACCESSION  CB401772
VERSION     CB401772.1 GI:30743499
KEYWORDS   EST.
SOURCE     Caenorhabditis elegans
ORGANISM   Caenorhabditis elegans
REFERENCE  1 (bases 1 to 565)
AUTHORS    Rebol,J., Vaglio,P., Rual,J.F., Lamesch,P., Martinez,M.,
Armstrong,C.M., Li,S., Jacotot,L., Bertin,N., Janky,R., Moore,T.,
Hudson,J.R., Hartley,J.L., Braach,M.A., Vandenhaute,J., Boulton,S.,
Endress,G.A., Jerna,S., Chevet,E., Papasotirooulos,V.,
Tolias,P.P., Ptacek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H.,
Doucette-Stamm,L., Hill,D.B. and Vidal,M.
C. elegans ORFome version 1.1: experimental verification of the
genome annotation and resource for proteome-scale protein
expression
Nat. Genet. (2003) In press
Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Sequence tag of Gateway entry clones. The primers used were
designed on the predicted protein encoding ORF. C. elegans ORFome
cloning project : Contact david_hill@dfci.harvard.edu or
marc-vidal@dfci.harvard.edu
POLYA=No.
Location/Qualifiers
1. .565

```

TITLE

```

JOURNAL    Nat. Genet. (2003) In press
COMMENT    Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Sequence tag of Gateway entry clones. The primers used were
designed on the predicted protein encoding ORF. C. elegans ORFome
cloning project : Contact david_hill@dfci.harvard.edu or
marc-vidal@dfci.harvard.edu
POLYA=No.
Location/Qualifiers
1. .565

```

ORIGIN

```

FEATURES
source
1. .565

```

/organism="Caenorhabditis elegans"  
/mol\_type="mRNA"  
/strain="N2"  
/db\_xref="taxon:6239"  
/sex="Hermaphrodite and male"  
/tissue\_type="whole animal"  
/dev\_stage="mixed stage"  
/clone\_lib="AD-wrmcDNA"

/notes="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pC86"

## ORIGIN

Alignment Scores: 2.57e-103 Length: 565  
Pred. No.: 999.00 Matches: 187  
Score: 99.5% Conservative: 0  
Percent Similarity: 99.5%  
Best Local Similarity: 99.5% Mismatches: 1  
Query Match: 18.7% Indels: 0  
DB: 4 Gaps: 0

US-10-645-746-3 (1-1020) x CB401772 (1-565)

```

QY 9 GluLysGlyPheTyrArgHisSerLeuAspProGluMetLysTrpLeuAlaArgProThr 28
DB 1 GAAAAAGGATTTTATCGTCAATCTCTCGATCCGAGATGAATGGCTTGGAGGCCCACT 60

QY 29 GlyLysCysAspGlyLysPheTyrGluLysValLeuLeuLeuValLeuLeuLeuLeu 48
DB 61 GGTAAATGCCAGCGCAAAATCTATGAGAAAGATGACTCTTTTGGTAAATGGTTCAAG 120

QY 49 PheSerSerLysIleTyrAspArgGluTyrTyrGluTyrGluValLysMetThrLysGlu 68
DB 121 TTCTCCAGCAAAATTTACGATCGGAATACCTACGAGTATGATGAGTGAATGACAGAGAA 180

QY 69 ValLeuAsnArgLysProGlyLysProPheProLysLysThrGluLeuProLeuProAsp 88
DB 181 GTATTGAATAGAAAACCCAGGAAACCTTTCCCAAAAAGACAGAAATTCATCCCGAT 240

QY 89 ArgAlaLysLeuPheTrpGlnHisLeuArgHisGluLysLysGlnThrAspPheIleLeu 108
DB 241 CGTGCAAACTCTTCGGCAACATCTTCGGCATGAGAAAGACAGACAGATTTTATTC 300

QY 109 GluAspTyrValPheAspGluLysAspThrValTyrSerValCysArgLeuAsnThrVal 128
DB 301 GAAGACTATGTTTTGATGAAAGGACACTGTTTATAGTCTTTGTCGACTGAACGCTGC 360

QY 129 ThrSerLysMetLeuValSerGluLysValValLysLysAspSerGluLysLysAspGlu 148
DB 361 ACATCAAAAATGCTGGTTTCGGGAAAGTAGTAAAAAAGGATTCGGAGAAAAAAGATGAA 420

QY 149 LysAspLeuGluLysLysIleLeuTyrThrMetIleLeuTyrTrpArgLysLysPheHis 168
DB 421 AAGGATTTGGAGAAAAAATCTTATACAAATGATCTTACCTATCGTAAAAAATTCAC 480

QY 169 LeuAsnPheSerArgGluAsnProGluLysAspGluAlaAsnArgSerTyrLysPhe 188
DB 481 CTGAACTTTAGTCGAGAAATCCGGAAGAAAGACGAGAGCGAATCGGAGTTACAAATTC 540

QY 189 LeuLysAsnValMetThrGlnLys 196
DB 541 CTGAAGAATGTTATGACCCAGAAA 564

```

## RESULT 10

BJ116268

## LOCUS

BJ116268 583 bp mRNA linear EST 23-JAN-2002  
Caenorhabditis elegans cDNA library, C. elegans L1 stage

## DEFINITION

Caenorhabditis elegans cDNA clone yk1201c08 5', mRNA sequence.

## ACCESSION

BJ116268

## VERSION

BJ116268.1

## KEYWORDS

EST.

## SOURCE

## ORGANISM

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 583)

Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.

and Sugano,S.

A complementary view of the C.elegans genome

Unpublished (2002)

Contact: Tadasi Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1..583

/organism="Caenorhabditis elegans"

/mol\_type="mRNA"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk1201c08"

/sex="hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="L1"

/clone\_lib="unpublished oligo-capped cDNA library, C.

elegans L1 stage"

ORIGIN

Alignment Scores: 7.77e-103 Length: 583

Pred. No.: 995.00 Matches: 193

Score: 99.5% Conservative: 0

Percent Similarity: 99.5% Mismatches: 1

Best Local Similarity: 99.5% Indels: 0

Query Match: 18.6% Gaps: 0

DB: 2

US-10-645-746-3 (1-1020) x BJ116268 (1-583)

QY 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysValValLys 140

DB 1 AGTGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60

QY 141 LysAspSerGluLysLysAspGluLysAspLeuGluLysLysLysLysLysLysLys 160

DB 61 AAGGATTCGAGAAAAAAGATGAAGAGATTCGAGAAAAAATCTTATACACATGATA 120

QY 161 LeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGluLysAspGlu 180

DB 121 CTTACCTATCGTAAAAAATTTTACCTGAATTCGAGAAAAATCCGAAAAAAGACGAA 180

QY 181 GluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysValArgTyrAla 200

DB 181 GAAGCAATCGGAGTTTACAAATTCCTGAAGATGTTATGACCCAGAAAGTTCGTCAC 240

QY 201 PropheValAsnGluGluLysValGlnPheAlaLysAsnPheValTyrAspAsnAsn 220

DB 241 CCTTTTGTACAGCGAGATTAAGTACAATTCGCAAAAATTTTGTGTACGATAAAT 300

QY 221 SerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGlnSerLeuGlu 240

DB 301 TCAATTCGCGAGTTCTCGAATCGTTTTCAGATCCAAACAGATTCGAACCAATCAATGAA 360

QY 241 ValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyLysGluLeuPheAsp 260

DB 361 GTAGCACCAAGAAATCGAAGCATGGTTTGGAAATTTACATTCGAATCAAGAAATTTGTT 420

QY 261 GlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAlaProLysMet 280

DB 421 GGTGAACCTGTGCTCAATTTTGCATTTGCGATTAACATTTCTACATGACCGCAATG 480

QY 281 SerLeuLeuAspTyrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 300



```

Db      481  TCTCTCTGGATATCTTCTCTCTAATGTGCGACCCCACTGCTGTAACGATGATGACGA 540
Qy      301  LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIle 314
Db      541  AAAGATCTTAAACAAACTGATGGCGGGAATAATGACCAATC 582

RESULT 11
CB398209/c
LOCUS   CB398209_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
DEFINITION
ACCESSION CB398209
VERSION   CB398209.1 GI:30739936
KEYWORDS EST.
SOURCE   Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE
AUTHORS Rebol,J., Vaglio,P., Rual,J.P., Lamesch,P., Martinez,M.,
Armstrong,C.M., Li,S., Jacotot,L., Bertin,N., Janky,R., Moore,T.,
Hudson,J.R., Hartley,J.L., Brasch,M.A., Vandenhaute,J., Boulton,S.,
Endress,G.A., Jenna,S., Chevet,E., Papasotiropoulos,V.,
Tollas,P.P., Placek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H.,
Doucette-Stamm,L., Hill,D.E. and Vidal,M.
TITLE    C. elegans ORFeome version 1.1: experimental verification of the
genome annotation and resource for proteome-scale protein
expression
JOURNAL Nat. Genet. (2003) In press
COMMENT  Contact: Vidal M
        Marc Vidal Laboratory
        Dana Farber Cancer Institute
        1 Jimmy Fund Way Smith 658, BOSTON, MA 02115, USA
        Tel: 617 632 5180
        Fax: 617 632 5739
        Email: Marc.Vidal@dfci.harvard.edu
        Sequence tag of Gateway entry clones. The primers used were
        designed on the predicted protein encoding ORF. C. elegans ORFeome
        cloning project : Contact david_hilledfci.harvard.edu or
        marc_vidal@dfci.harvard.edu
        POLYA=No.

FEATURES
source      Location/Qualifiers
1..554
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pFC86"

ORIGIN
Alignment Scores:
Pred. No.:      1.71e-101      Length:      554
Score:          983.00      Matches:    184
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    18.4%      Indels:     0
DB:             4      Gaps:        0

US-10-645-746-3 (1-1020) x CB398209 (1-554)

Qy      831  AspGlyGluAspProGluProLysThrThrPheIleValIleGlyLeuArgHisAsnThr 850
Db      553  GATGGAGAGATCCAGACCGCGAAGTACACGTTTCATTGTGATTCAGAAAGACACATACA 494
Qy      851  ArgLeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeuThrProAla 870

```

```

Db      493  CGATTGCTTCGAAGAAATGGAATAAGCAGTGGTCAATAAGATCTTACTCTCGCT 434
Qy      871  GluThrAspValAlaValAlaValLysGlnTrpGluGluAspMetLysGluSerLys 890
Db      433  GAAACAGATGTCGCTGTTGCTCTTTAAACAATGGGAGGAGATATGAAGAAGCAAA 374
Qy      891  GluThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLys 910
Db      373  GAAACTGGAATTTGTGAACCCATCATCCGGAACAACTGTGGATAAACTTATCGTTTCGAAA 314
Qy      911  TyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGly 930
Db      313  TACAAATTCGATTTTTTCTTGGCATCTCATCATGCTGTCCTTGGTACATCTCGTCCAGCA 254
Qy      931  HisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThr 950
Db      253  CATTACACTGTTATGTATGACGATTAAGGAATAGCAAGATGAAGTCTTATAAAATGACC 194
Qy      951  TyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValProVal 970
Db      193  TACGACACTGCTTTTCTCTCTAGATGTGCAAAACCATCTCGTTGCTGCTGTCGGTT 134
Qy      971  HisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990
Db      133  CATTATGCTCATTTATCATGTGAAAAGCGAAAGAGCTTTATCGAACTTTACAAGGAACAT 74
Qy      991  TyrIleGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnThr 1010
Db      73  TACATCGGTGACTATGCACGACCGACGACGAAATCGAAATGGAACATTTTCTCCAAACT 14
Qy      1011  AsnValLysTyr 1014
Db      13  AACGTGAAGTAC 2

RESULT 12
BJ105221
LOCUS   BJ105221
DEFINITION
ACCESSION BJ105221
VERSION   BJ105221.1 GI:18247891
KEYWORDS EST.
SOURCE   Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE 1 (bases 1 to 557)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE    A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT  Contact: Tadasu Shin-i
        Center For Genetic Resource Information
        National Institute of Genetics
        1111 Yata, Mishima, Shizuoka 411-8540, Japan
        Tel: 81-559-81-6856
        Fax: 81-559-81-6855
        Email: tshini@genes.nig.ac.jp.
        Location/Qualifiers
1..557
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1070f12"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"

ORIGIN

```



```

Alignment Scores:
Pred. No.: 5,13e-97 Length: 557
Score: 944.00 Matches: 184
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 99.5% Mismatches: 0
Query Match: 17.6% Indels: 0
DB: 2 Gaps: 0

US-10-645-746-3 (1-1020) x BJ105221 (1-557)

QY 675 IleAlaLeuYsIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTyrSerGlu 694
DB 1 ATTGCATTGAAATCAACGCGAATATTAGGAGTATTACCCAGAGCTTGACTGCTCAGAA 60

QY 695 IleAlaGluIleSerProGluGluGluGluGluGluGluGluGluGluGluGluGlu 714
DB 61 ATTGCAGATAATATCACCAGAGAAAGAAAGAACGAGAAACAAATGCCATTAATATGAT 120

QY 715 ValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAla 734
DB 121 GTTGGAAATTGATGTAATCTATCCAACTCTCCACAGTGGAAATTGATTTATCTATAGCGGCT 180

QY 735 ValValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGlu 754
DB 181 GTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAATATGATTGTGACTCAGAA 240

QY 755 GluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGlu 774
DB 241 GAAATGTCGCCGTGAGCGTGCAGTGGCTCATCGACGGGAAAGAACAGATATTTTGGAA 300

QY 775 AlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAspAsnArgAlaProAla 794
DB 301 GCAAGTTCTGGAATTGCTCAGAGAATTGCGAGAAACCAACGCAATCCAGACCCAGCG 360

QY 795 HisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAsp 814
DB 361 CATATTGATCTATCGAGACGGATTAGCGATTCGGAGATGCTACGTGTAGTCATGAT 420

QY 815 GluLeuArgSerLeuLysSerGluValLysGluPheMetSerGluArgAspGlyGluAsp 834
DB 421 GAGCTTCGATCTTTAAGAACGGAAGTAAACAAATTCATGTCGGAACGGGATCGAGAAGAT 480

QY 835 ProGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArg 854
DB 481 CCGAGCCGAAGTACAGCTTTCATTTGTGATTGAGAAAGACAAATACAGATTGCTTCGA 540

QY 855 ArgMetGluLysAsp 859
DB 541 AGAATCGAAGAAAGAT 555

RESULT 13
BJ127233
LOCUS
DEFINITION
BJ127233 unpublished oligo-capped cDNA library, EST 23-JAN-2002
Caenorhabditis elegans cDNA clone yk1342d04 5', mRNA sequence.
ACCESSION
BJ127233
VERSION
BJ127233.1 GI:18287390
KEYWORDS
EST.
SOURCE
Caenorhabditis elegans
ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 534)
Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome
Unpublished (2002)
Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..534
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1342d04"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"

FEATURES
source
1..534
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..534
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1342d04"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"

ORIGIN
Alignment Scores:
Pred. No.: 3,28e-95 Length: 534
Score: 928.00 Matches: 177
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 17.3% Indels: 0
DB: 2 Gaps: 0

US-10-645-746-3 (1-1020) x BJ127233 (1-534)

QY 363 LysValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrLeuPhe 382
DB 3 AAAGTCACTGGTAAATCGGACAGAGGAAGAAATGCAAAAAAGTAGTACGATACACTTGTTC 62

QY 383 LysIleTyrGluGluAsnLysPheIleGluPheProHisLeuProLeuValLysVal 402
DB 63 AAAATCTATGAGGAAACAAAAAATTCATTGAGTTTCCACCCTACCACTAGTCAAGATT 122

QY 403 LysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysPro 422
DB 123 AAAAGTGGACAAAGAAAGATACGCTGTACCAATGGAACATCTTGAAGTTTCATGAGAGCCA 182

QY 423 GlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr 442
DB 183 CAAAGATACAAAGATCGAATTGATCTGGTGTATGCAAGCAAGTTTCTAAGCGAGCTACA 242

QY 443 ArgLysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSer 462
DB 243 CGAAAACTTCACGACTACAAAGAAAAATACCTTAAAAATGCTGAAAGAAATTTGGATTCTCT 302

QY 463 SerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGlu 482
DB 303 TCTGAAGAGCTAAATTTGTTTGAAGATTTGGATTATGCTCCAAACTTCAGATGATCGAA 362

QY 483 CysProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLys 502
DB 363 TGTCCAGGAAGGTTTGAAGAGCCAAATGCTTGTGAATAGTGTAAATGAACAAATTTAA 422

QY 503 MetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGluLysGlu 522
DB 423 ATGACACCAAGTGTATGCTGGATTTCAAGAAAAACAATTTGAATGTTGGTTCCGAAAAAGAA 482

QY 523 LeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeu 539
DB 483 CTTTGTCTGTCTGTTTTTTGTAGTCAAGAAACAGCGGGAATCCATGCTTA 533

RESULT 14
BJ127087
LOCUS
DEFINITION
BJ127087 unpublished oligo-capped cDNA library, EST 23-JAN-2002
Caenorhabditis elegans cDNA clone yk1342d11 5', mRNA sequence.
ACCESSION
BJ127087
VERSION
BJ127087.1 GI:18287244
KEYWORDS
EST.
SOURCE
Caenorhabditis elegans
ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

```

```

REFERENCE 1 (bases 1 to 539)
AUTHORS Kohara,Y., Shin-I.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
Location/Qualifiers
1..539
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1342d11"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"

ORIGIN
Alignment Scores:
Pred. No.: 2,97e-93 Length: 539
Score: 911.00 Matches: 177
Percent Similarity: 98.3% Conservative: 0
Best Local Similarity: 98.3% Mismatches: 2
Query Match: 17.0% Indels: 1
DB: 2 Gaps: 0

US-10-645-746-3 (1-1020) x BJ127087 (1-539)

QY 306 LysLeuMetAlaGlyLysMetThrIleArgGlnAlaAArgProArgIleArgGlnLeu 325
DB 2 AAACGTATGCGGGAAAATGACATACAGACGCCGCGGCCAAGATTCGACAAATTA 61

QY 326 LeuGluAsnLeuLysLeuLysCysAlaGluValTTPAspAsnGluMetSerArgLeuThr 345
DB 62 TTGGAAAATTTGAAGCTGAATTCGCGAGAAGTTTGGGATAACGAAATGTCGAGATTGACA 121

QY 346 GluAqGHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThr 365
DB 122 GAACGACATCTGACATTTCTAGATTTGTGCGANGAAACTCTCTTTTATAAAGTCACT 181

QY 366 GlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyr 385
DB 182 GGTAAATCGGACAGAGAGAGAAATGCANAAANAGTAGCTACATTGTTCAAAATCTAT 241

QY 386 GluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGly 405
DB 242 GAGGAAAAACAAAAGTTTCAATTCAGTTTCCGCCACTACCACTANTCAAAGTTAAAGTGA 301

QY 406 AlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyr 425
DB 302 GCAAAAGAATACGCTGTACCAATGGAAACATCTTGAAGTTTCATGAGAAAGCCACAAAGATAC 361

QY 426 LysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysPro 445
DB 362 AGAATCGAATTGATCTGGTATGCAAGACAGATTTCATAAGCGAGCTACACGAAACCT 421

QY 446 HisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGlu 465
DB 422 CACGACTACAAAGAAAATACCCCTAAATAATGCTGAAAGAATTCGATTTCTCTTCTGAAGAG 481

QY 466 LeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIle-GluCysPro 484
DB 482 CTAAATTTTGTGAAAGATTGGATTATGCTCCAAACTTCAGATGATCCGAATGTCCA 539

```

RESULT 15

```

C62850
LOCUS C62850 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk296b10 5', mRNA sequence.
ACCESSION C62850
VERSION C62850.2 GI:56147596
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 378)
AUTHORS Kohara,Y., Mochashii,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 22, 1997 this sequence version replaced gi:2421555.
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..378
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk296b10"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="Yuji Kohara unpublished cDNA"

ORIGIN
Alignment Scores:
Pred. No.: 7,51e-66 Length: 378
Score: 670.00 Matches: 122
Percent Similarity: 98.4% Conservative: 0
Best Local Similarity: 98.4% Mismatches: 2
Query Match: 12.5% Indels: 0
DB: 4 Gaps: 0

US-10-645-746-3 (1-1020) x C62850 (1-378)

QY 1 MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspProGlu 20
DB 7 ATGTCTCGAATTTTCCGGAATTCGAAAAGGATTTATCGTCATCTCTCGATCCGGAG 66

QY 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysLysVal 40
DB 67 ATGAAATGGCTTCGAGGCCCCACTGGTANATCGACGGCAAAATTTATGAGAAAGATGA 126

QY 41 LeuLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGlu 60
DB 127 CTTCTTTTGGTAAATTTGGTTCAAGTTCTCCACCAAAATTTACATCGGGAATACACGAG 186

QY 61 TyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLys 80
DB 187 TATGAAGTGAAATGCAAGGAAGTNTTGAATAGAAAACCAAGGAAAACCTTTCCCAAAA 246

QY 81 LysThrGluIleProIleProAspArgAlaLysLeuPheTrpGluHisLeuArgHisGlu 100
DB 247 AAGACAGAAAATTCCAATTCCTCCGATTCCTGCAAACTCTTCTGCAACATCTTCGGCATGAG 306

QY 101 LysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAspThrValTyr 120
DB 307 AAGAGGACAGACAGATTTTATTTCTCGAAGACTATGTTTTTGTATGAAAANGACACTGTGTTAT 366

QY 121 SerValCysArg 124
|||||

```

Db 367 AGTGTTCGCA 378

Search completed: July 6, 2006, 04:44:16  
Job time : 9539 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 5, 2006, 22:12:21 ; Search time 1292 Seconds  
(without alignments)  
8256.610 Million cell updates/sec

Title: US-10-645-746-3

Perfect score: 5349

Sequence: 1 MSSNPFELKGRHSLDPE.....RHEHFLOQTNVYKPGMSFA 1020

Scoring table:

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US10645746/runat\_05072006\_110142\_5038/app\_query.fasta.1  
-DB=N Geneseq -QFMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs05p  
-USER=US10645746 @CGN\_1\_1\_1147 @runat\_05072006\_110142\_5038 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_8:\*

1:	Geneseqn1980s:*
2:	Geneseqn1990s:*
3:	Geneseqn2000s:*
4:	Geneseqn2001as:*
5:	Geneseqn2001bs:*
6:	Geneseqn2002as:*
7:	Geneseqn2002bs:*
8:	Geneseqn2003as:*
9:	Geneseqn2003bs:*
10:	Geneseqn2003cs:*
11:	Geneseqn2003ds:*
12:	Geneseqn2004as:*
13:	Geneseqn2004bs:*
14:	Geneseqn2005s:*
15:	Geneseqn2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5349	100.0	3227	4	AAS03283
2	4804.5	89.8	3709	4	AAS03282
3	604.5	11.3	2960	13	ADQ80698

4	577	10.8	2571	12	ADJ94703	Adj94703 Human euk
5	577	10.8	7478	10	ADB81486	ADB81486 DNA of th
6	577	10.8	7478	14	AEC01623	AEC01623 Human EIF
7	575.5	10.8	2910	9	ACD98411	ACD98411 A. thalia
8	575.5	10.8	2910	10	ADC46702	ADC46702 Thalesres
9	575.5	10.8	2910	12	ADI43620	ADI43620 Plant tra
10	575.5	10.8	2910	12	ADO02338	ADO02338 Thalesres
11	566	10.6	2580	12	ADJ94704	ADJ94704 Human euk
12	566	10.6	3011	6	ABQ99304	ABQ99304 Human cod
13	566	10.6	3580	6	ABS78717	ABS78717 Human cdn
14	565	10.6	3996	4	AHL14510	AHL14510 Human cdn
15	564.5	10.6	2955	6	ABSS1392	ABSS1392 CDNA enco
16	564.5	10.6	3486	4	ABL20457	ABL20457 Drosophil
17	564.5	10.6	3586	4	ABL20459	ABL20459 Drosophil
18	564.5	10.6	3743	4	ABL11513	ABL11513 Drosophil
19	556.5	10.4	2568	12	ADJ94706	ADJ94706 Human euk
20	556.5	10.4	2914	4	AAK53425	AAK53425 Human pol
21	556.5	10.4	2914	8	ABX13625	ABX13625 Human cyt
22	556.5	10.4	3050	4	AAH16058	AAH16058 Human cdn
23	556.5	10.4	3050	12	ADH77114	ADH77114 Human PAZ
24	555.5	10.4	3736	8	ABX34752	ABX34752 Human mdd
25	555	10.4	2772	12	ADJ94705	ADJ94705 Human euk
26	550.5	10.3	3307	14	ADM42124	ADM42124 Human euk
27	550.5	10.3	3325	8	ADSS5860	ADSS5860 Human nuc
28	530.5	9.9	2815	13	ADFS5695	ADFS5695 Human PRO
29	523.5	9.8	2960	13	ADQ80704	ADQ80704 Arabidops
30	523.5	9.8	2973	12	ADO61720	ADO61720 Transcrip
31	519.5	9.7	3008	12	ADO63642	ADO63642 Transcrip
32	519.5	9.7	8465	4	ABL20458	ABL20458 Drosophil
33	519.5	9.7	10149	4	ABL20456	ABL20456 Drosophil
34	519.5	9.7	12558	4	ABL11512	ABL11512 Drosophil
35	519	9.7	3120	13	ADQ80694	ADQ80694 Arabidops
36	519	9.7	3147	12	ADI43622	ADI43622 Plant tra
37	519	9.7	3326	3	AAC47079	AAC47079 Arabidops
38	515	9.6	3071	14	ADY65087	ADY65087 S. manson
39	510.5	9.5	3040	13	ADQ80700	ADQ80700 Arabidops
40	509	9.5	3195	3	AAC50992	AAC50992 Arabidops
41	501	9.4	2960	13	ADQ80696	ADQ80696 Arabidops
42	501	9.4	2967	3	AAC50603	AAC50603 Arabidops
43	501	9.4	2967	6	ABZ12627	ABZ12627 Arabidops
44	496.5	9.3	3718	13	ADX45492	ADX45492 Plant ful
45	495	9.3	2042	9	ADB17457	ADB17457 Corn post

#### ALIGNMENTS

##### RESULT 1

AAS03283  
ID AAS03283 standard; cDNA; 3227 BP.

XX  
AC AAS03283;

XX  
DT 07-SEP-2001 (first entry)

XX  
DE C. elegans cDNA encoding RNA interference pathway protein RDE-1.

XX  
KW RNA interference; RNAi; RDE-1; genetic interference; antibody; dsRNA;  
XX double-stranded RNA-dependent gene silencing; ss.  
XX Caenorhabditis elegans.

Key	Location/Qualifiers
5'UTR	1..20
FT	/*tag= a
FT	21..3083
FT	/*tag= b
FT	/product= "RDE-1"

XX  
PN WO200129058-A1.

XX  
PD 26-APR-2001.

XX  
PF 13-OCT-2000; 2000WO-US028470.



Db 1461 ATCGAATGTCAGGAAAGGTTTTGAAGAGCCAAATGCTTGTGAATAGTGTAAATGAACAA 1520  
 Qy 501 IleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAenValProGlu 520  
 Db 1521 ATTAATAATGACACCAAGTGAATTCGTGGATTTTCAGAAAAACAATGGAATGCTGTTCCCGAA 1580  
 Qy 521 LysGluLeuCysCysAlaValPheValValAenGluThrAlaGlyAenProCysLeuGlu 540  
 Db 1581 AAGAACTTTGCTGCTGTTTTTGTAGTCAACGAAACAGCGGAAATCCATGCTTAGAA 1640  
 Qy 541 GluAenAspValValLysPheThrThrGluLeuIleGlyGlyCysLysPheAtrGlyIle 560  
 Db 1641 GAGAAACGACGTTGTTAAAGTTCTACACCGAACTAATGCTGGTTCGAAGTTCCTGGGAATA 1700  
 Qy 561 ArgIleGlyAlaAenGluAenArgGlyAlaGlnSerIleMetThrAspAlaThrLysAen 580  
 Db 1701 CGAATTGCTGCCAATGAAACAGAGGAGCGCAATCTAATTATGTACGACGCGACGAAAAAT 1760  
 Qy 581 GluThrAlaPheThrLysAenCysThrLeuAenThrGlyIleGlyArgPheGluIleAla 600  
 Db 1761 GAATATGCCCTTACAAAAATTTGACACTAAATACCGGAATCGGTAGATTGGAATAGCC 1820  
 Qy 601 AlaThrGluAlaLysAenMetPheGluArgLeuProAspLysGluGlnLysValLeuMet 620  
 Db 1821 GCAACAGAGCGAAGATATGTTTGAACGCTTCCCGATTAAGAACAAAAAGCTTAATG 1880  
 Qy 621 PheIleIleSerLysArgGlnLeuAenAlaThrGlyPheValLysHisThrCysAsp 640  
 Db 1881 TTCAATTATCATTTCCAAACGACCAACTGAATGCTTACCGTTTGTGAAACATTATTGGCAT 1940  
 Qy 641 HisThrIleGlyValAlaAenGlnHisIleThrSerGluThrValThrLysAlaLeuAla 660  
 Db 1941 CACACCATCGGTAGCTAATACAGCATATTACTTCTGAAACAGTCAACAAAGCTTTGGCA 2000  
 Qy 661 SerLeuArgHisGluLysGlySerLysArgIlePheThrGlnIleAlaLeuLysIleAen 680  
 Db 2001 TCACCTAAGGCACGAGAAAGGATCAAAACGAATTTCTATCAATTCGATTGAAATCAAC 2060  
 Qy 681 AlaLysLeuGlyGlyIleAenGlnGluLeuAspThrSerGluIleAlaGluIleSerPro 700  
 Db 2061 GCGAAATTAGAGGTATTAAACAGGAGCTTGACTGGTCAGAAATTCGAGAAATATCACCA 2120  
 Qy 701 GluGluLysGluArgArgLysThrMetProLeuThrMetThrValGlyIleAspValThr 720  
 Db 2121 GAAGAAAAAGAAAGACGAAACAAATGCCAATTAATGCTATGATGTGTTGAAATGATGAAT 2180  
 Qy 721 HisProThrSerThrSerGlyIleAspThrSerIleAlaAlaValValAlaSerIleAen 740  
 Db 2181 CATCCAACTCTCTACAGTGGAAATGATTATTCTATAGCGGCTGTAGTAGCGATTCAT 2240  
 Qy 741 ProGlyGlyThrIleThrArgAenMetIleValThrGlnGluGluCysArgProGlyGlu 760  
 Db 2241 CCAGGTGGAACTATCTATCGAAATATGATTGTGACTCAAGAAAGATGTCGCCGGTGAG 2300  
 Qy 761 ArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeu 780  
 Db 2301 CGTGCACTGCTCATGACGCGGAAGAACAGATATTTTGGAAACAAAGTTTCGTGGAATG 2360  
 Qy 781 LeuArgGluPheAlaGluAenAenAspAsnArgAlaProAlaHisIleValValThrArg 800  
 Db 2361 CTCAGAGAAATTCGACAGAAACACAGCAATCGAGCACCGCATATTGTAGTCTATCGA 2420  
 Qy 801 AspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLys 820  
 Db 2421 GACGGAGTTAGCGAATTCGAGATGCTACGTTGTAGTCATGATGAGCTTCGATCTTTAAAA 2480  
 Qy 821 SerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLysThr 840  
 Db 2481 AGCGAAGTAAAAACAATTCATGTGCGAACCGGATGGAGAGATCCAGGCGCGAAGTACACG 2540  
 Qy 841 PheIleValIleGlnLysArgHisAenThrArgLeuLeuArgArgMetGluLysAspLys 860  
 Db 2541 TTCAATTGCTATCAGAAAGACACATACAGATTGCTTCCAGAAATGGAAAGAAATAG 2600

Qy 861 ProValValAenLysAspLeuThrProAlaGluThrAspValAlaValAlaAalaValLys 880  
 Db 2601 CCAGTGTCTAATAAGATCTTACTCTCTGCTGAAACAGATGCTGCTGTTGCTGCTGTAAA 2660  
 Qy 881 GlnTrpGluGluAspMetLysGluSerLysGluThrGlyIleValAenProSerSerGly 900  
 Db 2661 CAATGGGAGGAGGATATGAAAGAAAGCAAGAAACTGGAATTTGTGAACCCATCATCCGA 2720  
 Qy 901 ThrThrValAspLysLeuIleValSerLysThrLysPheAspPhePheLeuAlaSerHis 920  
 Db 2721 ACACTGTGGATAAACTTATCGTTTCGAATACAAATTCGAATTTTCTTGGCATCTCAT 2780  
 Qy 921 HisGlyValLeuGluThrSerArgProGlyHisThrThrValMetThrAspAspLysGly 940  
 Db 2781 CATGGTGTCTTGTACATCTCGTCCAGGACATTAACACTGTATGTATGATGACGATAAAGGA 2840  
 Qy 941 MetSerGlnAspGluValThrLysMetThrThrGlyLeuAlaPheLeuSerAlaArgCys 960  
 Db 2841 ATGAGCCAAAGATGAAGTCTATAAAATGACCTACGACTTGTCTTCTCTGCTAGATGT 2900  
 Qy 961 ArgLysProIleSerLeuProValProValHisThrAlaHisLeuSerCysGluLysAla 980  
 Db 2901 CGAAACCCATCTCGTGTCTGCTTCCGTTCAATATGCTATTTATCATGTGAANAAGCG 2960  
 Qy 981 LysGluLeuThrArgThrThrLysGluHisThrIleGlyAspThrAlaGlnProArgThr 1000  
 Db 2961 AAAGAGCTTTATCGAACTTACAAGGAACATTAACACTGTGACTATGCACAGCCAGGACT 3020  
 Qy 1001 ArgHisGluMetGluHisPheLeuGlnThrAsnValLysThrProGlyMetSerPheAla 1020  
 Db 3021 CGACAGAAATGGAACATTTTCTCCAAACTAACTGAAGTACCTGGAATGTCTGTTGCA 3080

RESULT 2  
 AAS03282  
 ID AAS03282 standard; DNA; 3709 BP.  
 XX  
 AC AAS03282;  
 XX  
 DT 07-SBP-2001 (first entry)  
 XX  
 DE C. elegans genomic DNA encoding RNA interference pathway protein RDE-1.  
 XX  
 KW RNA interference; RNAi; RDE-1; genetic interference; antibody; dsRNA;  
 KW double-stranded RNA-dependent gene silencing; ds.  
 XX  
 OS Caenorhabditis elegans.  
 XX  
 Key Location/Qualifiers  
 CDS 21..3582  
 FT /\*tag= a  
 FT /product= "RDE-1"  
 FT /note= "This sequence contains introns"  
 XX  
 PN WO200129058-A1.  
 XX  
 PD 26-APR-2001.  
 XX  
 PF 13-OCT-2000; 2000NO-US028470.  
 XX  
 PR 15-OCT-1999; 99US-0159776P.  
 PR 30-MAR-2000; 2000US-0193218P.  
 XX  
 PA (UYMA-) UNIV MASSACHUSETTS.  
 PA (CARN-) CARNegie INST WASHINGTON.  
 XX  
 PI Mello CC, Fire A, Tabara H, Grishok A;  
 XX  
 DR WPI; 2001-316239/33.  
 DR P-PSDB; AAU01856.  
 XX  
 PT Novel RNA interference pathway genes and their protein products involved  
 in mediation of genetic interference, useful for modulating and studying

PT regulation of RNA interference pathway.

XX Claim 1; Fig 5; 76pp; English.

XX The sequence is genomic DNA encoding the RNA interference (RNAi) pathway protein RDE-1. RDE-1 and RDE-4 are involved in the pathway mediating double-stranded RNA-dependent gene silencing (genetic interference). RDE-1 and RDE-4 protein is useful for preparing an RNAi agent, by incubating a dsRNA in the presence of the proteins. The prepared RNAi agents can be used as sequence-specific interfering agents for targeted genetic interferences. The nucleic acids are useful for studying the regulation of RNAi pathway and to generate knockout strains of animals such as C.elegans. RDE-1 and RDE-4 genes and their products are useful for modulating RNAi pathway activity. The polypeptides are useful for generating and testing antibodies specific for the polypeptides which are useful for studying the RNAi pathway in C.elegans and other organisms. RNAi pathway genes are useful for mediating specific processes, e.g. a gene that mediates dsRNA uptake by the cells is useful for transporting other RNAs into cells or for facilitating entry of agents such as drugs into cells. RNAi pathway mutant strains (rde-1, rde-4 mutants) are useful in genetic screens to identify additional RNAi pathway components

XX Sequence 3709 BP; 1308 A; 621 C; 741 G; 1039 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 3709  
Score: 4804.50 Matches: 1004  
Percent Similarity: 84.4% Conservative: 3  
Best Local Similarity: 84.2% Mismatches: 13  
Query Match: 89.8% Indels: 182  
DB: 4 Gaps: 10

US-10-645-746-3 (1-1020) x AAS03282 (1-3709)

QY 1 MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspPro--- 19  
DB 21 ATGTCTCGAAATTTTCCGGAATTGGAAAAAGGATTTATCGTCATCTCTCGATCCGGTA 80  
QY 20 -----GluMetIly 22  
DB 81 TGAATCAATTTATTAGCAGCTATTAAGATATATAAAGTTTGTATTAATTAATATTAGGAGATGAA 140  
QY 22 sTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysLysValLeuLe 42  
DB 141 ATGGCTTGGCGGCCACCTGGTAAATCGCAGCGCAATTTCTATGAGAAAGATCTTCT 200  
QY 42 uLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGluTyrG1 62  
DB 201 TTTGGTAAATTTGGTTCAAGTTCTCCAGCAAAATTTTACGATCGGGAATACTACGAGTATGA 260  
QY 62 uValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysTh 82  
DB 261 AGTGAATAATGACAAAAGGAAGTATTGAATAGAAAACCCAGAAAACCTTTCCCAAAAAGA- 319  
QY 82 rGluIleProfile-----ProAs 88  
DB 320 -GAAATTCGAATGTAAGTCTGTAAATTAGTCAAAACTAATTTTATTTTTCAGTCCCGA 378  
QY 88 pArgAlaLysLeuPheTrpGlnHisLeuArgHisGluLysLysGlnThrAspPheIleLe 108  
DB 379 TCGTGCMAAACCTCTTCTGGCAACATCTTCGGCATCGAAGAAGCAGACAGATTTTATTCT 438  
QY 108 uGluAspTyrValPheAspGluLysAspThrValTyrSerValCysArgLeuAsnThrVa 128  
DB 439 CGAAGCATATGTTTTTGTATGAAAAGGACATCTGTTATAGTGTGTTTGTGCACTGAACACTGT 498  
QY 128 lThrSerLysMetLeuValSerGluLysValLysLysAspSerGluLysLysAspG1 148  
DB 499 CACATCAAAAATGCTGGTTCCGAGAAAGTAGTAATAAAGGATTCGGAGAAAAGATGA 558  
QY 148 uLysAspLeuGluLysLysIleLeuTyrThrMetIleLeuThrTyrArgLysPheHi 168  
DB 559 AAGGATTTTGGAGAAAAAAATCTTATACAAATGATACTTACCTATCTCGTAAAAAATTTCA 618

QY 168 sLeuAsnPheSerArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyrLysph 188  
DB 619 CCGTGAATTTAGTTCGAGAAAATCCGGAAGAAAGACGAAAGCGAATTCGGAGTTTACAAAT 678  
QY 188 eLeuLys-----AsnValMetTh 194  
DB 679 CTTGAAGGTTTATGAAAAACACGCAATTATACAAACAATTAGCTTTTCAGAAATGTTATGAC 738  
QY 194 rGlnLysValArgTyrAlaProPheValAsnGluGluLysVal----- 209  
DB 739 CCAGAAAGTTCGCTACGCGCCTTTTGTGAACGAGGAGATTAAGTGTGAGTTTGCATAAAT 798  
QY 210 -----GlnPheAlaLysAsnPheVa 216  
DB 799 AATAATAATAATCACTCACTCAATTTATATATATTTTAAAGACAATTCGCGAAAAATTTGT 858  
QY 216 lTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheG1 236  
DB 859 GTACGATTAATTAATTCATTTCTCGAGTTCCTGAATCGTTTCACGATCCAAACAGATTCGA 918  
QY 236 uGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyIleLy 256  
DB 919 ACAATCATTAGAAGTAGCACCAGAAATCGAAGCATGTTTGAATTTTACATTGGAATCAA 978  
QY 256 sGluLeuPheAspGlyGluProValLeuAsnPheAlaIle----- 269  
DB 979 AGAATTTGTTTCGATGTAACCTGTCTCAATTTTCCAAAG-TAAGTTTGGAAACTCGCAT 1037  
QY 270 -----ValAspLysLeuPheTyrAsnAlaProLysMe 280  
DB 1038 AAAAAATCATGTGATTTTGTGAAGTTGTCGATAAACTATTCTACATGCAACCGAANAAT 1097  
QY 280 tSerLeuLeuAspTyrLeuLeuLeuLeuValAspProGlnSerCysAsnAspAspValAr 300  
DB 1098 GTCTCTTCTGGATATCTTCTCTTAATTTGTCGACCCCGCTGCTGTAAAGCATGATGACG 1157  
QY 300 GlyAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgPr 320  
DB 1158 AAAAGATCTTAAACAAAAACTGATGCGCGGAAAAATGCAATCAGCAAGCCGCGCGGCC 1217  
QY 320 oArgIleArgGlnLeuLeuGluAsnLysLeuLysCysAlaGluValTrpAspAsnG1 340  
DB 1218 AAGAATTCGACAAATTTTGGAAAAATTCGAAGCTGAAATGCGCAGAGAATTTGGGATAACAA 1277  
QY 340 u-----MetSerArgLeuTh 345  
DB 1278 ATGTTAGTTTAAATTTATTTCAAAACAATTAATATACAAATTTGATTTTTCAGGTCGAGATTGAC 1337  
QY 345 rGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValTh 365  
DB 1338 AGAACGATCTGACATTTCTAGATTTGTGCGAGGAAAACTCTCTCTGTTTATTAAGTCAC 1397  
QY 365 rGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTy 385  
DB 1398 TGGTAAATCGCAGACAGAGCAAGAAATGCAAAAAAAGTACGATACTACTATTCTTCAAAATCTA 1457  
QY 385 rGluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerG1 405  
DB 1458 TGAGGAAAAACAAAAGTTTCATTGAGTTTCCCCACCTACCACTAGTCAAGTTTAAAGTGG 1517  
QY 405 YAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTy 425  
DB 1518 AGCAAAAGAAATACGCTGTATCCATGGAACATCTTGAGTTCTATGAGAGCCCAAGATA 1577  
QY 425 rLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysPr 445  
DB 1578 CAAGAAATCGAATTTGATCG-GTGA TGCAAGACAAAGTTTCTTAAAGCGAGCTTACACGAAAAACC 1636  
QY 445 oHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluG1 465  
DB 1637 TCACGACTACAAAGAAAAATCCCTTAAAAAATGCTGAAAAA-TTGGATTTCTCTTCTGAAGA 1695



QY 465 uLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlu 485  
 DB 1696 GCTAAATTTTGTGAAAGATTGGATATGCTCCAAACTTCAGATGATGCAATGTCAGG 1755  
 QY 485 YLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPr 505  
 DB 1756 AAGGTTTTGAAGAGCCCAATGCTTGTGAATAGTGAATGAACAAATTAATGACACC 1815  
 QY 505 ovalIleArgGlyPheGlnGluLysGlnLeuAsnValValProGluLysGluLeuCysCy 525  
 DB 1816 AGTGATTCGTGGATTTCAAGAAACAAATTTGAATGTGTTCCCGAAAGAACTTTCGTC 1875  
 QY 525 sAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnValVa 545  
 DB 1876 TCGTGTITTTGTAGTCAACGAAACAGCGGAAATCCATGCTTGAAGAGAACGACGTTGT 1935  
 QY 545 l-----LysPheTyrThrGluLeuI 552  
 DB 1936 GTAAGTGTITTTCTACGPAGATTATTCGAAATATTTTCAGTAAGTTCTACCGNACTAA 1995  
 QY 552 lGlyGlyCysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgIleValaGlns 572  
 DB 1996 TTGGTGGTTTCCAAAGTTCCGTGGAAATCCGAATTTGTCGAATGAAACACAGAGGAGCGCAAT 2055  
 QY 572 erIleMetTyrAspAlaThrLysAsnGlu----- 581  
 DB 2056 CTATTATGTACGACGCGCAAAATATGAAATATGTCGCTAAGTTTCAGAAATTTGAAAGTTT 2115  
 QY 582 -----TyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyValArgP 597  
 DB 2116 TTAATATCATATTTACAG-TTCTACAAAAATTTGTACACTTAATACCGGAATCGGTAGAT 2174  
 QY 597 heGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnL 617  
 DB 2175 TTGAATAGCGCAACAGAGCGCAAGAAATATGTTGAACGCTTCCCGATAAAGAACAAA 2234  
 QY 617 ysValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysH 637  
 DB 2235 AGTC-TTAATGTTCATTATCATTTCCAAACGACCACTGAATGCTTACGGTTTTTGTGAAC 2293  
 QY 637 isTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrL 657  
 DB 2294 ATTATTTCCGATCACACCATCGGTGTAGCTTAATCAGCATATTACTTCTGAACAGTACAA 2353  
 QY 657 ysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaL 677  
 DB 2354 AAGCTTTGGCATCACTAAGGACGACGAGAAAGGATCAAAACGAATTTCTATCAAAATTCAT 2413  
 QY 677 eulysIleAsnAlaLysLeuGlyIleAsnGlnGluLeuAspTyrSerGluIleAlaG 697  
 DB 2414 TGAATAATCAACGCGAAATTTAGGAGGTATTAACCAGGAGCTTGACTGGTCAGAAATTCAG 2473  
 QY 697 luIleSerProGluGluLysGlu-ArgArgLysThrMetProLeuThrMetTyrValGly 716  
 DB 2474 AAATATCACCAAGAAAGAAAGAAAGACGGAACCAATGCCATTACTATGATGTGGA 2533  
 QY 717 IleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValVal 736  
 DB 2534 ATTGATGTAACCATCCAACTCC-TACAGTGGAAATGATTTATTTCTATAGCGCTGTAGTA 2592  
 QY 737 AlaSerIleAsnProGlyThrIleTyrArgAsnMetIleValThrGlnGluGluCys 756  
 DB 2593 GCGAGTATCAATCCAGTGGAACTATCTATCGAAATATGATTTGTGACTCAAGAAGAAATGT 2652  
 QY 757 ArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLys 776  
 DB 2653 CGTCCCGGTGAGGTGCGAGTGGCTCATGACGCGGAAGAACAGATATTTTGGAGC-AAA 2711  
 QY 777 PheValLysLeuLeuArgGluPheAlaGlu----- 786  
 DB 2712 TTTCGTGAAATTTGCTCAGAGAAATTCGACAGAGTTCGAGTTGCTTCAGTATTTTAAAGATCTC 2771  
 QY 787 -----AsnAsnAspAsnArgAlaProAlaHisIle 796

DB 2772 TGGATTTTAAATTTTGTAAACTTTTCAGAACCAACGACAATCGACCGCATAT 2831  
 QY 796 eValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLe 816  
 DB 2832 TGTAGTCTATCAGACGCGAGTTAGCGATTTCGGAGATGCTACGTGTTAGTCATGATGACT 2891  
 QY 816 uArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGlu 836  
 DB 2892 TCGATCTTTAAAGCGAGTAACAAATTCATGTCGGAACGGATCGAGAAATCCAGA 2951  
 QY 836 uProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgArgMe 856  
 DB 2952 GCCGAAGTACACGTTTCATTGTGATTTCAGAAAAGACACAAATACAGATTGCTTCGAAGAAT 3011  
 QY 856 tGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAlaVa 876  
 DB 3012 GGAAGAAAGATAAGCCAGT-GTCAATAAAGATCTTACTCTCTGCTGAAACAGATGTCGTGT 3070  
 QY 876 lAlaAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyIleValAs 896  
 DB 3071 TGCTGCTGTTAAACAATGGGAGGAGGATATGAAGAAAGCAAGAAACTGGAAATTGTGAA 3130  
 QY 896 nProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPhePh 916  
 DB 3131 CCCATCATCCGGAACAACGTGTGATAAATTCATGTTTCGAAATACAAATTCGAATTTTTT 3190  
 QY 916 eLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTy 936  
 DB 3191 CTTGGCATCTCATGTTGTCCTTGGTACATCTCTGTCAGGACATTTACACTGTTATGTA 3250  
 QY 936 rAspAspLysGlyMetSerGlnAspGluValTyr----- 947  
 DB 3251 TGACGATAAAGGAATGAGCCAGATGAAGTCTATGTAAGCGTTTGAATAGCAGTTAGCG 3310  
 QY 948 -----LysMetThrT 951  
 DB 3311 ATTTTAGGATTTTGTAAATCCGCATATAGTTATTATATAAAAAATGTTTCAGAAAAATGACCT 3370  
 QY 951 YrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValH 971  
 DB 3371 ACGGACTTGTCTTCTCTCTGATAGTGTGAAACCCATCTCGTTGCTGTTCCGGTTC 3430  
 QY 971 isTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 991  
 DB 3431 ATTATGCTCATTTATCATGTGAAAGCGAAAGAGCTTTATCGAACTTACAAGAACATT 3490  
 QY 991 YrIleGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnThra 1011  
 DB 3491 ACATCGGTGACTATGCACAGCCACGAGCTCGACACGAAATGGAACATTTTCTCCAACTA 3550  
 QY 1011 snValLysTyrProGlyMetSerPheAla 1020  
 DB 3551 ACGTGAAGTACCTCGAATGTGCTTCGCA 3579

## RESULT 3

ADQ80698

ID ADQ80698 standard; DNA; 2960 BP.

XX AC ADQ80698;

XX XX

DT 21-OCT-2004 (first entry)

XX DE Arabidopsis thaliana TFL1-binding protein coding sequence #7.

XX XX TFL1-binding protein; plant growth control; biotechnology;

XX XX fishing industry; screening; gene; ds.

XX OS Arabidopsis thaliana.

XX XX Location/Qualifiers

FT CDS 1..2958

FT /\*tag= a







```

Db      1581  TGTGAGTGAACGTCGAGATACACTCTTGGGAATGGCTACCGAGTGTGCAGGT 1640
Qy      652  rGlutThrValThrIysAlaLeuAlaSerLeuArgHisGlySerIysArgIlePh 672
Db      1641  GAAGAACGTCGTCAAGACC-----TCACCTCAGACTCT 1673
Qy      672  eTyrGlnIleAlaLeuIysIleAenAlaLysLeuGlyGlyIleAenGlnGluLeuAspTr 692
Db      1674  GTCCAACTCTGCCTCAAGATCAATGTCAATCTTGGTGGCATTAAACAATCCTTA----- 1728
Qy      692  pSerGluIleAlaGluIleSerProGluGluLysGluArgArgLysThrMetProLeuTh 712
Db      1729  -----GTCCACACACAGCGCTCGCCGTTTTCACAGCCAGG-- 1767
Qy      712  rMetTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIl 732
Db      1768  -ATATTCTGGGAGCAGATGTTACACACCCCGCAGCGGGATGGGAAAAAACCTTCTAT 1826
Qy      732  eAlaAlaValValAlaSerIleAenProGlyGlyThrIleTyrArgAenMetIleValTh 752
Db      1827  CACAGCAGTGTAGCGATGATGGATGCCACCCAGCGCGATACCTGTCTACTGTGCGGT 1886
Qy      752  rGlnGluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIl 772
Db      1887  ACAG-----CGACCACGCGAAGAT 1907
Qy      772  eLeuGlu-----AlaLysPheValLys---LeuLeuArgGluPheAlaGluAenAenAs 789
Db      1908  CATTGAAGACTTGTCTACATGTCGTGAGCTCTCTCATCTCAATTCATCAAGTCCACCCG 1967
Qy      789  pAenArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLe 809
Db      1968  TTCCAAG---CCTACCCGCGATCATCTTCTACCGAGATGGGTCCTGAAGCCAGCTACC 2024
Qy      809  uArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGl 829
Db      2025  CGAGTACTCCACTATGAGTACTGGCCATTCGTGATGCTGCATCAACTG----- 2076
Qy      829  uArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHisAs 849
Db      2077  -----GAAAGAGACTACCAGCTGGGATCACTTATATTGTGTGTCAGAAACGCCATCA 2129
Qy      849  nThrArgLeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeuThrPr 869
Db      2130  CACCCGCTTTTC---TGTGCTGACAGAATGAGCGAATTTGGGAAGAGTGGTAACATCCC 2186
Qy      869  oAlaGluThrAspValAlaValAlaLaValLysGlnTrpGluGluAspMetLysGluSe 889
Db      2187  AGCT----- 2190
Qy      889  rLysGluThrGlyIleValAenProSerSerGlyThrThrValAspLysLeuIleValSe 909
Db      2191  -----GGGACCACAGTGGACACCAACATCACCACA 2219
Qy      909  rLysTyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGlyThrSerArgPr 929
Db      2220  CCAATTTGAGTTTGACTTCTATCTGTGCAGCCAGCAGGATCCAGGSCACCAAGCCGACC 2279
Qy      929  oGlyHisThrValMetTyrAspAspLysGlyMetSerGluAspGluValTyrLysMe 949
Db      2280  ATCCATTATATGTTCTTTGGGATGACAACCGTTTCACAGCAGATGAGTCCAGATCCT 2339
Qy      949  tThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValPr 969
Db      2340  GACGTACCAGCTGTGCCACACTTAGGTACGATGCACAGCTCTGTCTCTATCCAGCACC 2399
Qy      969  ovalHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGl 989
Db      2400  TGCTACTATGCGCGCTGTGGCTTTCCGGGACGCA---TACCACCTGTGTGGACCAAGGA 2456
Qy      989  uHisTyrIleGlyAsp 994

```

```

Db      2457  GCATGACAGTGGAGAG 2472
RESULT 5
ADB81486
ID      ADB81486 standard; DNA; 7478 BP.
XX
AC      ADB81486;
XX
DT      04-DEC-2003 (first entry)
XX
DE      DNA of the human eukaryotic translation initiation factor 2C 1 mRNA.
XX
KW      gene; ds; human; eukaryotic translation initiation factor 2C 1; EIF2C1;
KW      chromosome 1p34-35; Co-EIF2C; EIF2C; Golgi ER protein 95kDa; GERp95; Q99;
KW      gene therapy; hyperproliferative disorder;
KW      familial hypercholesterolaemia; cancer; polycystic kidney disease;
KW      cystic fibrosis; progeroid syndrome; cytostatic; antilipaemic.
XX
OS      Homo sapiens.
XX
Key      Location/Qualifiers
CDS      214..2787
          /*tag= a
          /product= "EIF2C1 protein"
FT
FT
FT
XX
PN      WO2003040321-A2.
XX
XX      15-MAY-2003.
XX
XX      04-NOV-2002; 2002WO-US035324.
XX
XX      08-NOV-2001; 2001US-00007078.
XX
XX      (ISIS-) ISIS PHARM INC.
XX
XX      Ward DT, Watt AT;
XX
XX      WFI; 2003-449448/42.
XX
XX      P-PSDB; ADB81572.
XX
XX      New compound, having a sequence targeted to a nucleic acid encoding human
XX      collapsin response mediator protein 2, useful for preparing a composition
XX      for treating hypercholesterolemia or hyperproliferative disorder, e.g.,
XX      cancer.
XX
XX      Example 13; Page 82-87; 120pp; English.
XX
XX      This invention relates to novel antisense oligonucleotides that modulate
XX      the expression of human eukaryotic translation initiation factor 2C 1
XX      (EIF2C1). EIF2C1 is located on chromosome 1p34-35, and is also known as
XX      Co-EIF2C, EIF2C, Golgi ER protein 95kDa, GERp95 and Q99. It is an
XX      intracellular membrane associated protein thought to be involved in
XX      cellular differentiation, such that altered expression of EIF2C1 can
XX      affect cell growth, morphology and tumorigenicity. Accordingly,
XX      antisense oligonucleotides that inhibit the expression of EIF2C1 in cells
XX      or tissues can be used in gene therapy to treat various conditions
XX      including hyperproliferative disorders, familial hypercholesterolaemia
XX      and cancer, as well as polycystic kidney disease, cystic fibrosis and
XX      progeroid syndrome. As such, the oligos of the present invention can be
XX      described as having cytostatic and antilipaemic activities. This
XX      polynucleotide sequence is the DNA of the human eukaryotic translation
XX      initiation factor 2C 1 (EIF2C1) mRNA of the invention.
XX
XX      Sequence 7478 BP; 1750 A; 1863 C; 1895 G; 1970 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 6,06e-43 Length: 7478
Score: 577.00 Matches: 237
Percent Similarity: 40.2% Conservative: 173
Best Local Similarity: 23.2% Mismatches: 370
Query Match: 10.8% Indels: 243
DB: 10 Gaps: 38

```

US-10-645-746-3 (1-1020) x ADB81486 (1-7478)

QY 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGlu----- 37  
DB 209 ATGGGATGAAA-GCGGGACCTCGGAGCAGCTGCGGGCGCTTACCTGCCCCCCTGCAG 267  
QY 38 -----LysLysValLeuLeu 42  
DB 268 CAGGTGTTCCAGGCACCTCGCGGCTGGCATTGGCAGCTGTGGGGAACCAATCAAGCTC 327  
QY 43 LeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgLysTyrGluTyrGlu 62  
DB 328 CTGGCCAAATTACTTTGAGGTGGACATCCCTAAGATCGACCTGTGACCCTACGAGGTGGAC 387  
QY 63 ValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysThr 82  
DB 388 ATCAAGCCGGATAAG-----TGTCCTCCCGTAGAGTC 417  
QY 83 GluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGluLysLys 102  
DB 418 AACCGGAAGTGGTGAA-----TACATGGTCAGCATTTCAAGCCT 459  
QY 103 GlnThrAspPheIleLeuGluAspTyr-----ValPheAspGluLysAspThrValTyr 120  
DB 460 CAG-----ATCTTTGGTGTATCGAAGCCTGTGTATGATGGAAGAAGAACATTATAC 510  
QY 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysVal----- 138  
DB 511 -----ACTGTACAGCAGCTGCCCATTTGGCAACAGCGGTGCGACTTT 552  
QY 139 ---ValLysLysAspSerGluLysLysAspGluLysLysLysLysLysLys 157  
DB 553 GAGGTGACAACTCCCTGGGAGAGGAAGAT---CGAATCTTTAAGTCTCCATCAAGTGG 609  
QY 158 ThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGlu 177  
DB 610 CTAGCCATTGTGAGCTGGCGAATGCTGCATGAGCGCTGTGCAGCGCCAGATCCCTGTT 669  
QY 178 LysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysVal 197  
DB 670 CCTTGGAGTCTGTGCAAGCCCTGGATGTGGCATGAGGCATCTGGCATCC-----ATG 723  
QY 198 ArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyr 217  
DB 724 AGGTACACCCCTGTG----- 738  
QY 218 AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln 237  
DB 739 ---GGCGGCTCTCTTCTCAGCGCTGAGGGCTACTACACCCG----- 780  
QY 238 SerLeuGluValAlaProArgIleGluAlaTyrPheGlyIleTyrIleGlyLysGlu 257  
DB 781 -----CTGGGGGTGGCGCGAAGTCTGTTCTGGCTTTCACCAGTCTGTGGCCCT 831  
QY 258 LeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAla 277  
DB 832 GCCATG---TGGAGATGATGCTCAACATGATGTCTCAGCCACTGCTTTTATAAGGCA 888  
QY 278 ProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSerCysAsnAsp 297  
DB 889 -----CAGCCAGTGTGAGTTGATGTGAGGTGTGACATCATCAGGAACATAGATAG 942  
QY 298 AspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAla 317  
DB 943 CAGCCCAAG-----CCCTCAGCGACTCT 966  
QY 318 AlaAtProArgIleArgGlnLeuLeuGluAsnLeuLysLysLys-----CysAla 334  
DB 967 CAGCGGTTTCGCTTCCACCAAGGAGATCAAGGGCTGAAGGTGGAATCACCACCTGTGGA 1026  
QY 335 GluValTrpAspAsn-----GluMetSerArgLeuThrGluArgHisLeuThr 350  
DB 1027 CAGATGAAGAGGAAGTACCGCGTGTGTAATTTACCCTGCGCCTGCTAGCCATCAGACA 1086

QY 351 PheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLysSerAspArg 370  
DB 1087 TTCCCTTTACAGCTGAGAGTGGACAGACTGTG----- 1119  
QY 371 GlyArgAsnAlaLysLysTyrThrLeuPheLysIleTyrGluGluAsnLysLys 390  
DB 1120 -----GAGTGCACAGTGGCACAGTATTTCAAGCAGAAATATAAC 1158  
QY 391 Phe-----IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTyr 409  
DB 1159 CTTACAGTCAAGTATATCCCATCTGCCCTGTACAAAGTTTGGCCAGGAACAAAAGCATACC 1218  
QY 410 AlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArgIle 429  
DB 1219 TACCTTCCCTAGAGTCTGTAAACAT---GTGGCTGGCAGCGCTGTATTAAAGAGCTG 1275  
QY 430 AspLeuValMetGlnAspLysPheLysLysArgAlaThrArgLysProHisAspTyrLys 449  
DB 1276 ACCGCAACACAGACCTCGACCATGATAAAGGCCACAGCTAGATCCGCTCCAGACAGACAG 1335  
QY 450 GluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheVal 469  
DB 1336 GAGGAGATCAGTCGCTGTATGAAGATGCCAGCTACAACTTAGATCC-----TACATC 1389  
QY 470 GluArgPheGlyLysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys 489  
DB 1390 CAGGAATTTGGATCAAAAGTGAAGATGACATGACGAGGTGACAGGGGAGTGTGCGG 1449  
QY 490 GluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro----- 505  
DB 1450 GCGCCATCTTGCAGTACGCGCGCGGAACCGGGCATTTGCCACACCCCAATCAGGTGTC 1509  
QY 506 ---ValIleArgGlyPheGlnGluLysGlnLeu---AsnValValProGluLysGluLeu 523  
DB 1510 TGGGCAATGCGGGG-----AAACAGTTCTCAATGGATTTGAGATCAAAAGTCTGG 1560  
QY 524 CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp 543  
DB 1561 GCCATCGCTGTTCGACCCCAAAA-----CAGTGTGAGAAGAG----- 1602  
QY 544 ValValLysPheTyrThrGluLeuIleGly-----GlyCysLysPhe 557  
DB 1603 GTGCTCAAGAACTTCACAGACAGCTGCGGAAGATTTTCAAGGATGCGGGGATGCTATC 1662  
QY 557 eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAl 577  
DB 1663 CAGGTCACCTTGTTCGCAAAATATGCACAGGGGGCA-GACAGGCTG----- 1710  
QY 577 aThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPhe 597  
DB 1710 ----- 1710  
QY 597 eGluIleAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLys 617  
DB 1711 -----GAGCTATGTTCCGGCATCTCAAGAACACCTTACTCAGG 1748  
QY 617 sValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHis 637  
DB 1749 GCTCAGCTCATATTGTTCATCTCTCCAGGGAAGACCGCGGTGTATGCTGAGGTGAACG 1808  
QY 637 sTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLys 657  
DB 1809 TGTCGAGATACACTCTTGGGAATGGCTACGAGTGTGTGAGGTGAAGAACGTGTGCTCAA 1868  
QY 657 sAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLe 677  
DB 1869 GACC-----TCACCTCAGACTCTGTCCAACTCTCTGCCT 1901  
QY 677 uLysIleAsnAlaLysLeuGlyIleAsnGlnGluLeuAspTrpSerGluIleAlaGln 697  
DB 1902 CAAGATCAATGTCAAACTTGGTGGCATTAACAACATCTCTA----- 1941



QY 697 uileSerProGluGluLysGluArgArgLysThrMetProLeuThrMetTyrValGlyI 717  
 Db 1942 ----GTCCACACAGCGCTCTGCGTGTTCACAGCCAGT---ATATTCTGGGAGC 1994  
 QY 717 eAspValThrHisProThrSerTyrSerGlyLeuAspTyrSerIleAlaValValAl 737  
 Db 1995 AGATGTTTACACACCCCGCAGCGGGATGGAAAAACCTTCTATCACAGCGTGTAGG 2054  
 QY 737 aSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGluCysAr 757  
 Db 2055 CAGTATGGATGCCACCCAGCCGATCTGTGCTACTGTGCGGTACAG----- 2103  
 QY 757 gProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGlu-----AL 775  
 Db 2104 -----CGACCAACGGCAAGATCATTTGAAGACTGTGTC 2135  
 QY 775 aLysPheValLys---LeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAl 794  
 Db 2136 CTACATGTTGGTGAGTCTCTCATCCCAATTTACAGTCCACCGTTCAG-----CTAC 2192  
 QY 794 aHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAs 814  
 Db 2193 CCGCATCATCTTCTACCGAGATGGGTGCTGAAAGGCCAGCTACCCAGATACCTCCATTA 2252  
 QY 814 pGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAs 834  
 Db 2253 TGAGCTACTGGCCATCTCGTGATGCTGCATCAACTG-----GAAAGGA 2297  
 QY 834 pProGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuAr 854  
 Db 2298 CTACAGCCTGGGATCATTTATATTGTTGGTGCAAGAACGCATCACACCGCTTC-- 2355  
 QY 854 gArgMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspVa 874  
 Db 2356 -TGTGCTGACAGAATGAGCGAATTGGGAAGAGTGTAAACATCCAGCT----- 2403  
 QY 874 lAlaValAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyI 894  
 Db 2403 ----- 2403  
 QY 894 eValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAs 914  
 Db 2404 -----GGGACCACAGTGGACACCAATCACCCACCATTTGATTTGA 2447  
 QY 914 pPhePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrVa 934  
 Db 2448 CTTCTATCTGTGCAGCCACGAGCATCCAGGCGCACCGCCATCCATTACTATGT 2507  
 QY 934 lMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAl 954  
 Db 2508 TCTTTGGATGACAACCGTTTTCACAGATGAGCTCCAGATCCTGACGTACAGCTGTG 2567  
 QY 954 aPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHi 974  
 Db 2568 CCACATTAGTACGACAGCGCTGTCTTATCCAGACCTGCTCTACTATGCCCCG 2627  
 QY 974 sLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAs 994  
 Db 2628 CCTGTGGCTTTCGGGCAAG---TACCACCTGTGTGGACAAGAGCATGACAGTGGAGA 2684  
 QY 994 p 994  
 Db 2685 G 2685  
 RESULT 6  
 ID AEC01623  
 XX AEC01623 standard; DNA; 7478 BP.  
 AC AEC01623;  
 XX  
 DT 20-OCT-2005 (first entry)  
 XX  
 DE Human EIF2C1 DNA, SEQ ID NO: 3.

XX Cancer; cytostatic; neoplasm; hypercholesterolemia; antilipemic;  
 KW metabolic disorder; gene; ds; EIF protein kinase family;  
 KW eukaryotic translation initiation factor protein kinase family;  
 KW chromosome 1.  
 XX Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 CDS 214..2787  
 FT /\*tag= a  
 FT /product= "Human EIF2C1 protein"  
 XX  
 PN US2005182015-A1.  
 XX 18-AUG-2005.  
 PD 14-JAN-2005; 2005US-00035669.  
 PF 23-FEB-2001; 2001US-00793807.  
 PR 12-SEP-2001; 2001US-00954679.  
 PR 13-SEP-2001; 2001US-00953611.  
 PR 08-NOV-2001; 2001US-00007078.  
 XX (WARD/) WARD D T.  
 PA (WATT/) WATT A T.  
 XX  
 PI Ward DT, Watt AT;  
 DR WPI; 2005-563220/57.  
 DR P-PSDB; AEC01709.  
 DR REFSEQ; NM\_012199.2.  
 XX  
 PT New antisense oligonucleotides which inhibits expression of eukaryotic  
 PT initiation factor 2C1, useful for modulating RNA interference and  
 PT treating a disease or condition characterized by hypercholesterolemia, or  
 PT cancer.  
 XX  
 PS Example 13; SEQ ID NO 3; 67pp; English.  
 CC The present invention relates to an antisense oligonucleotide which  
 CC specifically hybridizes with the polynucleotide encoding eukaryotic  
 CC translation initiation factor 2C1 (EIF2C1; also known as Co-EIF-2C,  
 CC eIF2C, golgi ER protein 95kDa, GERP95 and Q99) and inhibits its  
 CC expression. The invention is useful for treating hypercholesterolemia and  
 CC hyperproliferative disorder such as cancer. The present sequence is the  
 CC human EIF2C1 DNA which is located on chromosome 1p34-p35.  
 XX  
 SQ Sequence 7478 BP; 1750 A; 1863 C; 1895 G; 1970 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 6.06e-43 Length: 7478  
 Score: 577.00 Matches: 237  
 Percent Similarity: 40.2% Conservative: 173  
 Best Local Similarity: 23.2% Mismatches: 370  
 Query Match: 10.8% Indels: 243  
 DB: 14 Gaps: 38  
 US-10-645-746-3 (1-1020) x AEC01623 (1-7478)  
 QY 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGlu----- 37  
 Db 209 ATGGGATGGAA-GCGGGAGCCCTCGGAGCAGCTGCGGCGCTTACCTGCCCCCTCGAG 267  
 QY 38 -----LysLysValLeuLeu 42  
 Db 268 CAGGTGTTCCAGGCACCTCGCGGCGCTGCGCATTTGGCAGCTGTGGGGAACCAATCAAGCTC 327  
 QY 43 LeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGluTyrGlu 62  
 Db 328 CTGGCCAATTAATTCTTGAGGTGGACATCCCTAAAGATCGACGTGTACCACTACGAGGTGAC 387  
 QY 63 ValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysThr 82









```

Db      2197 GAAGAAATTATTACAGGACCTGTATTAAGCTGGTTCCAGGATCCACCAACGTGGGCTAGTCCAC 2256
Qy      766 GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla 785
Db      2257 -----TCTGGTTTGTATAGGGAACATTTTCATAGCATTCAGGAGAGCT----- 2298
Qy      786 GluAsnAsnAspAsnArgAlaProAlaHisIleValValTyArgAspGlyValSerAsp 805
Db      2299 -----ACAGCCAGATACCTCAAGAGGATCATCTTCTATCGTACGGAGTAGGCGAA 2349
Qy      806 SerGluMetLeuArgValSerHisAspGluLeuLeuArgSerLeuLysSerGluValLysGln 825
Db      2350 GGGCAGTTAGTACAGGTTCTCTACATGAGATGATCTGCTATCCGACAGGCTTGTAACTCT 2409
Qy      826 PheMetSerGluArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGln 845
Db      2410 CTCGAA-----GAGAAATTATGTTCTCTCGTGTACTTTCGTGATGTGTCAG 2454
Qy      846 LysArgHisAsnThrArgLeuLeuArgMetGluLysAspLysProValValAsnLys 865
Db      2455 AAACGTCCACACACACAGCTTTGTTCCCTGAGCAACACCGG-----AATCGT 2499
Qy      866 AspleuThrProAlaGluThrAspValAlaValAlaAlaValLysGlnTrpGluLysAsp 885
Db      2500 GATATGACT----- 2508
Qy      886 MetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLys 905
Db      2509 -----GATAGAGTGGCAATATTCACCA-----GGTACTGCTGTGACACT 2550
Qy      906 LeuIleValSerLysTyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGly 925
Db      2551 AAAATCTGTACCCCTAATGAATTCGACTTCTATTGAAACGCCATGCTGTATTCAGGGA 2610
Qy      926 ThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGlu 945
Db      2611 ACAAGCAGCGCGCACATACCATGTTACTTCTCGATGAGACGGTTTCACCGCTGATCAG 2670
Qy      946 ValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSer 965
Db      2671 TTGCAAATGCTCACAACAACACCTCTGCTACACGTATCGGAGGTGTCACAAATCTGTCTCA 2730
Qy      966 LeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArg 985
Db      2731 ATTGTGCCACAGCTACTACGCTCACTTGGCTGCTATCGCTGCC-----CGC 2778
Qy      986 ThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThrArg 1001
Db      2779 TACTACATGGAGATGAGATGTCTGATGGAGTTCCGACGAGTCCAGG 2826

RESULT 8
ID      ADC46702
XX      ADC46702 standard; DNA; 2910 BP.
AC      ADC46702;
XX      18-DEC-2003 (first entry)
DT      XX
DE      Thalecress transcription factor-like DNA G1149.
KW      Thalecress; transcription factor-like protein; ds; seed trait;
KW      transgenic; gene; plant size; stress tolerance; yield;
KW      disease resistance; plant.
XX      Arabidopsis thaliana.
XX      OS
XX      US2003093837-A1.
XX      PN
XX      15-MAY-2003.
XX      PD
XX      01-NOV-2002; 2002US-00286264.
XX      PF
XX      23-MAR-1999; 99US-0125814P.
XX      PR

```

22-MAR-2000; 2000US-00533030.

(KEDD/) KEDDIE J.  
 (RIEC/) RIECHMANN J L.  
 (RATC/) RATCLIFFE O.  
 (ZHAN/) ZHANG J.  
 (JIAN/) JIANG C.  
 (PINE/) PINEDA O.  
 (HEAR/) HEARD J.  
 (YUGG/) YU G.  
 (ADAM/) ADAM L.  
 (BROU/) BROUN P.  
 (REUB/) REUBER L.  
 (PILG/) PILGRIM M.  
 (SAMA/) SAMAHA R.

Keddle J, Riechmann JL, Ratcliffe O, Zhang J, Jiang C, Pineda O;  
 Heard J, Yu G, Adam L, Broun P, Reuber L, Pilgrim M, Samaha R;  
 WPI; 2003-765498/72.  
 P-PSDB; ADC46703.

Novel transgenic plant having recombinant polynucleotide encoding  
 polypeptide that alters trait of transgenic plant when compared with same  
 trait of another plant lacking recombinant polynucleotide.

Disclosure; SEQ ID NO 101; 165pp; English.

The invention relates to a transgenic plant having recombinant  
 polynucleotide (II) encoding polypeptide comprising at least 6  
 consecutive amino acids of a sequence chosen from the protein sequence  
 appearing as ADC46603 - ADC46749 (every second sequence), where  
 recombinant polynucleotide alters a trait of the seed transgenic plant  
 when compared with same trait of another plant lacking recombinant  
 polynucleotide. The proteins are transcription factor-like proteins. Also  
 included are altering (M1) a trait associated with seed (comprising:  
 transforming a plant with (II); selecting the transformed plants; and  
 identifying a transformed plant with seed having altered trait); altering  
 (M2) the expression levels of at least one gene of a plant (involving  
 transforming the plant with (II) and selecting the transformed plant),  
 altering (M3) a trait associated with a plant's seed (comprising:  
 transforming the plant with a recombinant polynucleotide comprising a  
 nucleotide sequence comprising least 18 consecutive nucleotides of a  
 sequence appearing as ADC46750 - ADC46766 and selecting the transformed  
 plant) altering (M4) a plant's trait (involving providing a database  
 sequence, comparing the database sequence with a polypeptide or a  
 polynucleotide chosen as detailed above, selecting a database sequence  
 that needs selected sequence criteria and transforming a database  
 sequence in the plant) and altering a plant's trait (involving providing  
 a test polynucleotide, hybridising the test polynucleotide with a  
 polynucleotide detailed above at low stringency and transforming the  
 hybridising test polynucleotide in a plant to alter a trait of the  
 plant). The method (M1) is useful for altering a trait associated with  
 seed. The method (M2) is useful for altering the expressing levels of at  
 least one gene of a plant. The method (M3) is useful for altering a trait  
 associated with a plant's seed. The method (M4) is useful for altering a trait  
 of a plant's trait. The method (M4) is useful for altering a plant's trait  
 such as seed or plant size, stress tolerance, yield or disease  
 resistance. The present sequence encodes a transcription factor-like  
 protein/seed trait altering protein of the invention.

Sequence 2910 BP; 825 A; 590 C; 724 G; 771 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	2-24e-43	Length:	2910
Score:	575.50	Matches:	237
Percent Similarity:	39.0%	Conservative:	159
Best Local Similarity:	23.3%	Mismatches:	337
Query Match:	10.8%	Indels:	283
DB:	10	Gaps:	41

US-10-645-746-3 (1-1020) x ADC46702 (1-2910)









Db 2017 GATGCTATTAGAGAAACATACCTCTTATTACTGATCGTCCACCATCATCATCGGTGCT 2076  
 Qy 718 AspValThrHisProThrSerTySerGlyIleAspTySerIleAlaValAlaValAla 737  
 Db 2077 GATGTGACTCACCCACAGCCCTGGAGAGACTCAAGTCTCTTATTGCTGCTGTTGTGCC 2136  
 Qy 738 SerIleAsn---ProGlyGlyThrIleTyArgAsnMetIleValThrGln----- 753  
 Db 2137 TCTATGAGCTGCGCTGAGATAAACAATACCGAGATTGGTTTCTGCTCAAGCTCATAGG 2196  
 Qy 754 GluGlu-----CysArgProGlyGluArgAlaValAlaHis 765  
 Db 2197 GAAGAAATTATTTCAGGACCTGTATAAGCTGTTCCAGGATCCACAACTGGCTAGTCCAC 2256  
 Qy 766 GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla 785  
 Db 2257 -----TCTGGTTTGATAAGGGAACATTTTCATAGCATTCAGGAGAGCT----- 2298  
 Qy 786 GluAsnAsnAspAsnArgAlaProAlaHisIleValValTyArgAspGlyValSerAsp 805  
 Db 2299 -----ACAGGCCAGATACCTCAAGAGATCATCTTCTATCGTGACGAGTAAGCGNA 2349  
 Qy 806 SerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGln 825  
 Db 2350 GGGCAGTTTAGTCAGGTTCTGCTACATGATGATGCTGATCCCAAGGCTTGTAATCT 2409  
 Qy 826 PheMetSerGluArgAspGlyGluAspProGluProLysTyThrPheIleValIleGln 845  
 Db 2410 CTCGAA-----GAGAAATTATGTTCTCTGCTGTTACTTTTCGTGATTGTCCAG 2454  
 Qy 846 LysArgHisAsnThrArgLeuLeuArgArgMetGluLysAspLysProValValAsnLys 865  
 Db 2455 AAACGTCCACACACACGCTTTGTTCCCTCGACCAACACACGCG-----AATCGT 2499  
 Qy 866 AspLeuThrProAlaGluThrAspValAlaValAlaLysGlnTTPGluGluAsp 885  
 Db 2500 GATATGACT----- 2508  
 Qy 886 MetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLys 905  
 Db 2509 -----GATAGAGTGGCAATATTCAACCA-----GGTACTGCTGGGACACT 2550  
 Qy 906 LeuIleValSerLysTyLysPheAspPhePheLeuAlaSerHisHisGlyValLeuGly 925  
 Db 2551 AAATCTGTCCCTCAATGAATTCGACTTCTATTGAACGACCATGCTGTTATTCAGGGA 2610  
 Qy 926 ThrSerArgProGlyHisTyThrValMetTyAspAspLysGlyMetSerGlnAspGlu 945  
 Db 2611 ACAGCAGGCGCGCACATTACCATGTTCTCTCGATGAGAACCGTTTCCCGCTGATCAG 2670  
 Qy 946 ValTyLysMetThrTyGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSer 965  
 Db 2671 TTGCAAAATGCTCACAAACACCTCTGCTACACGATGCGAGGTGTACAAATCTGTGTCA 2730  
 Qy 966 LeuProValProValHisTyThrAlaHisLeuSerCysGluLysAlaLysGluLeuTyArg 985  
 Db 2731 ATTGTGCCACCGCTACTACGCTCACTTGGCTGCTTCGTCGCTCC-----CGC 2778  
 Qy 986 ThrTyLysGluHisTyIleGlyAspTyThrAlaGlnProArgThrArg 1001  
 Db 2779 TACTACATGAGAGTGAATGCTGTGATGGAGGTTCGAGCGGTCCTCAGG 2826

RESULT 10

ADO02338

ID ADO02338 standard; cDNA; 2910 BP.

XX AC ADO02338;

XX DT 01-JUL-2004 (first entry)

XX DE Thalearess transcription factor cDNA #376.

XX KW Thalearess; transcription factor; ss; gene; plant; transgenic;

KW abiotic stress; cold tolerance; heat tolerance; drought; osmotic stress;  
 KW phosphate limitation; potassium limitation; nitrogen limitation;  
 KW hormone sensitivity; disease resistance; sugar sensing; seed germination;  
 KW flowering; inflorescence architectural change;  
 KW meristem cell differentiation; phyllotaxy; apical dominance;  
 KW trichome development; seed development; premature senescence;  
 KW delayed senescence; lethality; necrosis; plant size; leaf morphology;  
 KW seed morphology; secondary metabolism; light response; shade avoidance.

OS Arabidopsis thaliana.

XX US2004045049-A1.

XX 04-MAR-2004.

XX 10-APR-2003; 2003US-00412699.

XX 13-SEP-1999; 99US-00394519.

XX 21-JAN-2000; 2000US-00489376.

XX 17-FEB-2000; 2000US-00506720.

XX 22-MAR-2000; 2000US-00532591.

XX 22-MAR-2000; 2000US-00533029.

XX 22-MAR-2000; 2000US-00533030.

XX 22-MAR-2000; 2000US-00533392.

XX 06-APR-2000; 2000US-00533648.

XX 16-NOV-2000; 2000US-00713994.

XX 27-MAR-2001; 2001US-00819142.

XX 17-APR-2001; 2001US-00837444.

XX 30-JAN-2002; 2002US-00958131.

XX 14-JUN-2002; 2002US-00171468.

XX 09-AUG-2002; 2002US-00225066.

XX 09-AUG-2002; 2002US-00225067.

XX 17-DEC-2002; 2002US-0434166P.

XX 25-FEB-2003; 2003US-00374780.

(ZHAN/) ZHANG J.

(FROM/) FROMM M E.

(HEAR/) HEARD J E.

(RIEC/) RIECHMANN J L.

(ADAM/) ADAM L J.

(BROU/) BROUN P E.

(PINE/) PINEDA O.

(REUB/) REUBER T L.

(KEDD/) KEDDIE J S.

(YUGG/) YU G.

(JIAN/) JIANG C.

(SAMA/) SAMAHA R S.

(PILG/) PILGRIM M L.

(CREE/) CREELMAN R A.

(DUBE/) DUBELL A N.

(RATC/) RATCLIFFE O.

(KUMI/) KUMIMOTO R.

(SHER/) SHERMAN B K.

Zhang J, Fromm ME, Heard JE, Riechmann JL, Adam LJ, Broun PE;

Pineda O, Reuber TL, Keddie JS, Yu G, Jiang C, Samaha RS;

Pilgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R;

Sherman BK;

WPI; 2004-225755/21.

P-PSDB; ADO02339.

XX New transgenic plant, useful in developing phenotypes with altered or  
 XX improved characteristics or traits.

XX Claim 1; SEQ ID NO 751; 213pp; English.

XX The invention relates to a transgenic plant comprises a recombinant  
 CC polynucleotide having a polynucleotide sequence or its complementary  
 CC sequence comprising a sequence encoding a polypeptide, that initiates  
 CC transcription (i.e. a transcription factor) from Arabidopsis, Soybean,





QY 443 ArgLysProHisAspTyrLysGluAenThrLeuLysMetLeuLysGluLeuAspPheSer 462  
DB 1477 CAACGACCCCTGAT---AGAGAGAACTCGATCAAAAACCTTGGTTGCAAAAATAATTAC 1533  
QY 463 SerGluGluLeuAenPheValGluArgPheGlyLeuCysSerLysLeuGluMetIleGlu 482  
DB 1534 AATGATCATCTGAC-----AGGAGTTTGGGATGTCAGTGACTACCCCACTGCTCG 1587  
QY 483 CysProGlyLysValLeuLysGluProMetLeuValAenSerValAenGluGlnIleLys 502  
DB 1588 ATTGAAGCTCGTGACTTCCCCACCGATGTTGAAGTACCATGACATGCTGAAGAGAA 1647  
QY 503 MetThrProValIleArgGlyPheGlnGluLysGlnLeuAenValProGluLysGlu 522  
DB 1648 ATGGTAATCCAAAGGCTAGGA-----CAGTGAACATGATGTTGACAGAAA--- 1692  
QY 523 LeuCysCysAlaValPheValValAenGluThrAlaGlyAenProCysLeuGluGluAen 542  
DB 1692 ----- 1692  
QY 543 AspValValLysPheTyrThrGluLeuIleGlyCysLysPheArgGlyIleArgIle 562  
DB 1693 -----ATGTTAATGGACAAA-----GTCACTTCT 1719  
QY 563 GlyAlaAenGluAenArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAenGluTyr 582  
DB 1720 TGGACTTGGCAATTTAAGCCTCAACCTGCTATTCCGTTTCTCTCTGTCCTCCCTGAACAT 1779  
QY 583 AlaPheTyrLysAenCysThrLeuAenThrGlyIleGlyArgPheGluIleAlaIaThr 602  
DB 1780 -----ATTGAGGAAGCTTCTCTC 1797  
QY 603 GluAlaLysAenMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPheIle 622  
DB 1798 GAT-----ATCCAAAGAGGACCTGCTCCAA-----CTGTTGATT 1836  
QY 623 IleIleSerLysArgGlnLeuAenAlaTyrGlyPheValLysHisTyrCysAspHisThr 642  
DB 1837 GTAATATTGCTGATGTGACTGATCATATGGAATAATAAAGGATCTGTGAACAGAA 1896  
QY 643 IleGlyValAlaAenGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeu 662  
DB 1897 TTGGGATGTCTCTCAGTGTGCCAACCTAGACAAAGTTAATAAATCTC----- 1944  
QY 663 ArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAenAlaLys 682  
DB 1945 -----AACAGCAGTACATCGAAATGTTGCTTGAAGATCAATGTCAG 1989  
QY 683 LeuGlyGlyIleAenGlnLeuAenTrpSerGluIleAlaGluIleSerProGluGlu 702  
DB 1990 ACTGGGGGAAGGAACACTGTTCTTAAT----- 2016  
QY 703 LysGluArgArgLysThrMetProLeu-----ThrMetTyrValGlyIle 717  
DB 2017 GATGCTATTAGAAGAACATACCTCTATTACTGATGTCCTCAACCATCATCATGGGTCT 2076  
QY 718 AspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValAla 737  
DB 2077 GATGTGACTCACCCACAGCTGGAGAGGACTCAAGTCTTCTATTGCTGCTGTGTGGCC 2136  
QY 738 SerIleAen---ProGlyGlyThrIleTyrArgAenMetIleValThrGln----- 753  
DB 2137 TCTATGGACTGGCTGAGATAAACAAATACCGAGGATGGTTTCTGCTCAAGCTCATAGG 2196  
QY 754 GluGlu-----CysArgProGlyGluArgAlaValAlaHis 765  
DB 2197 GAAGAAATTATTCCAGGACCTGTATAAGCTGGTTCCAGGATCCACACGTGGGCTAGTCCAC 2256  
QY 766 GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla 785  
DB 2257 -----TCTGTTGTATAGGGAACATTTTATAGCATTCAGGAGACT----- 2298

QY 786 GluAenAenAspAenArgAlaProAlaHisIleValValTyrArgAspGlyValSerAsp 805  
DB 2299 -----ACAGGCCAGATACCTCAAGAGGATCATCTCTATCTGTCGAGGAGTAAGCGAA 2349  
QY 806 SerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGln 825  
DB 2350 GGGCAGTTTATGTCAGGTTCTGCTACATGATGATGACTCTATCCGCAAGGCTTGTAACTCT 2409  
QY 826 PheMetSerGluArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGln 845  
DB 2410 CTCGAA-----GAGATTATGTTCTCTGCTGTTACTTTCGTGATTGTCAG 2454  
QY 846 LysArgHisAenThrArgLeuLeuArgMetGluLysAspLysProValValAenLys 865  
DB 2455 AAACGTCACACACACGTTGTTCCCTCGACCAACACCGG-----AATCGT 2499  
QY 866 AspLeuThrProAlaGluThrAspValAlaAlaAlaValLysGlnTrpGluGluAsp 885  
DB 2500 GATATGACT----- 2508  
QY 886 MetLysGluSerLysGluThrGlyIleValAenProSerSerGlyThrThrValAspLys 905  
DB 2509 -----GATAGAGTGGCAATATTCACCA-----GGTACTGCTGGACACT 2550  
QY 906 LeuIleValSerLysTyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGly 925  
DB 2551 AAAATCTGTCCACCTAATGAATTCGACTTCTATTGAACAGCCATGCTGTTATTTCAGGA 2610  
QY 926 ThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGlu 945  
DB 2611 ACAAGCAGGCGCGACATACCATGACTTCTCGATGAGAACCGTTTCACCGCTGATCAG 2670  
QY 946 ValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSer 965  
DB 2671 TTCGAAATGCTCACAACCAACCTCTGCTACACGATGCGAGGTGTACAAAATCTGTGTCA 2730  
QY 966 LeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArg 985  
DB 2731 ATTTGTCGCCACGCTACTACGCTCACTTGGCTGCATTCCGTGCC-----CGC 2778  
QY 986 ThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThrArg 1001  
DB 2779 TACTACATGAGAGTGAGATGTCTGATGGAGGTTCGAGCAGGTCCAGG 2826  
RESULT 11  
ID ADJ94704  
ID ADJ94704 standard; cDNA; 2580 BP.  
XX ADJ94704;  
XX 06-MAY-2004 (first entry)  
XX Human eukaryotic translation initiation factor 2C2 (eIF2C2) cDNA.  
XX cytostatic; antiinflammatory; virucide; immunosuppressive; tumour;  
XX inflammatory; infectious disease; viral infection; degenerative;  
XX autoimmune; gene therapy; Argonaute family;  
XX eukaryotic translation initiation factor 2C2; eIF2C2; human; ss; gene.  
XX Homo sapiens.  
XX WO2004007718-A2.  
XX 22-JAN-2004.  
XX 10-JUL-2003; 2003WO-EP007516.  
XX 10-JUL-2002; 2002EP-00015532.  
XX 23-AUG-2002; 2002EP-00018906.  
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX Tuschl T, Martinez J, Patkaniowska A, Urlaub H, Luehrmann R;





CC polynucleotides and proteins are useful for preventing, treating or  
 CC ameliorating disorders involving aberrant protein expression or  
 CC biological activity, e.g. haematopoietic disorders, central/peripheral  
 CC nervous system diseases, mechanical and traumatic disorders, non-healing  
 CC wounds, immune deficiencies and disorders, infectious diseases caused by  
 CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
 CC reactions and conditions, coagulation disorders, or cancer. The  
 CC polynucleotide sequences of the invention were assembled from ESTs  
 CC isolated mainly by sequencing by hybridisation, and in some cases,  
 CC sequences obtained from one or more public databases. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 3011 BP; 732 A; 877 C; 805 G; 597 T; 0 U; 0 Other;

Alignment Scores:      1.84e-42      Length:      3011  
 Pred. No.:      566.00      Matches:      239  
 Score:      40.0%      Conservatives:      159  
 Percent Similarity:      24.0%      Mismatches:      385  
 Best Local Similarity:      10.6%      Indels:      214  
 Query Match:      6      Gaps:      38  
 DB:

US-10-645-746-3 (1-1020) x ABQ99304 (1-3011)

Qy 19 ProGluMetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLys 38  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 198 CCATCCCAAGGATATGCTTCAAGCCTCCACCTAGACCGCAGCTTGGGACCTCCGGGAGA 257  
 Qy 39 LysValLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyr 58  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 258 ACATCAAAATTACAGCCCAATTTCTCGAATGGACATCCCAAAATTGACATCTATCAT 317  
 Qy 59 TyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPhe 78  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 318 TATGAATTGGATATCAAGCCAGCAGAG-----TGC 347  
 Qy 79 ProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArg 98  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 348 CCGAGGAGATTAAACAGGGAATCGTGGAAACATG-----GTCCAG 389  
 Qy 99 HisGluLysGlnThrAspPheIleLeuGluAspTyr-----ValPheAspGluLys 116  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 390 CACTTTAAACACAG-----ATCTTGGGATCGGAAGCCCGTGTTGACGGCAGG 440  
 Qy 117 AspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGlu 136  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 441 AAGAATCTATACACAGCCATGCCCTTCGATTGGGAGGACACAGGTG-----GAG 491  
 Qy 137 LysValValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysLysLeu 156  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 492 CTGGAGGTACGCTGCGAGCTTGCAGGCGTTACAGATGCACATTCAGGGCGGTGCCATC 545  
 Qy 157 TyrThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnPro 176  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 546 AAGTGGGTCTCTGCGTGCAGCTTGCAGGCGTTACAGATGCACATTCAGGGCGGTGCC 605  
 Qy 177 GluLysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGln--- 195  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 606 AGCGTCCCTTTTGAG-----ACGATCCAGGCCCTGGACGTGTCATGAGGCATG 656  
 Qy 196 ----LysValArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsn 214  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 657 CCATCCATGAGGTACACCCCGTG----- 680  
 Qy 215 PheValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArg 234  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 681 -----GGCGCTCTCTTCTACCGCGTCCGAAGGCTGCTCTAACCT----- 722  
 Qy 235 PheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGly 254  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 723 -----CTTGGCGGGGCCGAGAAGTGTGTTGGCTTTCATCAGTCC 764

Qy 255 IleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPhe 274  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 765 GTCCCGCTTCTCTC---TGGAAATATGTGTAATATGTGTGTCACAAACAGCGTT 821  
 Qy 275 TyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSer 294  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 822 TACAAGGA-----CAGCCAGTATCAGATTTGTTTGAAGTTTGTGATTTTAAAGT 875  
 Qy 295 CysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIle 314  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 876 ATTGAAGAACA-----CAAAACCTCTG 899  
 Qy 315 ArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLys----- 332  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 900 ACAGATTCCTCAAGGGTAAAGTTTACAAAGAAATTAAGGTTCAAGGTGAGATAACG 959  
 Qy 333 ---CysAlaGluValTrpAspAsn-----GluMetSerArgLeuThrGluArg 347  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 960 CACTGTGGCAGATCAAGAGNAGTACCGTGTCTGCAATGTGACCGCGCGCCGCACT 1019  
 Qy 348 HisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLys 367  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1020 CACCAAAACATTC---CCGCTGCAGCAGGAGCGGCGACAGCGTGGAGTGACG----- 1070  
 Qy 368 SerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGluGlu 387  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1071 -----GTGCCCATGAT-----TTCAG-----GACAG 1094  
 Qy 388 AsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLys 407  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1095 CACAAGTTGGTTCCTCCGCTACCCCACTCCCATGTTTACAAGTCGACAGGAGCAGAAA 1154  
 Qy 408 GluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsn 427  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1155 CACACTACCTTCCCTCGAGGTCTGTAACTTGTGGCAGGACAAAGATGTATTAAAAA 1214  
 Qy 428 ArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr---ArgLysProHis 446  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1215 TTAACGGAC-----AATCAGACCTCAACCATGATCAGAGCGACTGTAGTGGCGGCC 1268  
 Qy 447 AspTyrLysGluAsnThrLeuLysMetLysLysGluLeuAspPheSerSerGluLeu 466  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1269 GATCGCAAGAAGAGATTAGCAAAATTGATCGCAAGTGCAGATTTTCAACACAGATCCA--- 1325  
 Qy 467 AsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLys 486  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1326 ---TACGTCCGTGAATTTGGAATCATGTGTCAAAGATGATGACAGACGTGACTGGCGG 1382  
 Qy 487 ValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrProVal 506  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1383 GTGTGCGAGCGCCCTTCCATCTCTACGGGGCAGGAATAAGCTATTTCGACCCCTGTCT 1442  
 Qy 507 -----IleArgGlyPheGlnLysGlnLeuAsnVal---ValProGlyLysGluLeu 523  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1443 CAGGCGCTCGGACATCGGAACAAGCAGTTTCCACACGGGCATCGAGATCAAGGTGTGG 1502  
 Qy 524 CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp 543  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1503 GCCATTGCGTGTCTCGCCCCCAG-----CGCAGTGCACGGAAGTCCAT--- 1547  
 Qy 544 ValValLysPheTyrThrGluLeuIle-----GlyGly-CysLysPhe 557  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1548 ---CTGAAGTCTTTCACAGAGCAGCTCAGAAAGATCTCGAGAGAGCGCGCATGCCATC 1604  
 Qy 557 eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIle-----MetTy 575  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1605 CAGGCCAGCGCTGCTTCTGCAAAATACGCG-CAGGGGGCGGACAGCGTGGAGCCCATGTT 1663  
 Qy 575 rAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGln 595  
 |||    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::    ::  
 Db 1664 CCGCACCTGAAGAACACGTATGCG----- 1688

Qy 595 YArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysG1 615  
Db 1688 ----- 1688  
Qy 615 uGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAenAlaTyrGlyPheVa 635  
Db 1689 ----GGCTGCAGCTGGTGGTGGTCACTCTCCCGCAGACGCCCGGTAGCGCGAGGT 1744  
Qy 635 LysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrVa 655  
Db 1745 CAAGCGCTGGGAGACACGGTCTGGGGATGCCACGCGTGGTGCAGATGAAGAAGCT 1804  
Qy 655 lThrLysAlaLeuAlaSerLeuAArgHisGluLysGlySerLysArgIlePheTyrGlnI1 675  
Db 1805 G-----CAGAGACCCAGCCACAGACCTGTCCAACCT 1837  
Qy 675 eAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTrpSerGluI1 695  
Db 1838 CTGGCTGAAGATCAACGTCAAGCTGGGAGCGGTGAACAACATCTG----- 1883  
Qy 695 eAlaGluIleSerProGluGluLysGluArgLysThrMetProLeuThrMetTyrVa 715  
Db 1884 -----CTGCCCGCAGGCGCGCGGTGTTCAGCAGCGCGTC---ATCTTCT 1930  
Qy 715 lGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaIaVa 735  
Db 1931 GGGAGCAGAGTCACTCACCCCGCGGGGATGGGAAGCCCTCCATTCGCGCGT 1990  
Qy 735 lValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluI1 755  
Db 1991 GGTGGGCAGCATGAGACCCACCCATCTGCTGCGCCACCGTGGTGCAGCAG-- 2048  
Qy 755 uCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAl 775  
Db 2049 -----CACCGCAGCAGATCATACAAGACCTGGCGCG 2080  
Qy 775 aLysPheValLysLeuLeuArgGluPheAlaGluAenAsnAspAsnArgAlaProAlaHi 795  
Db 2081 CATGGTCCGCGAGCTCTCATCCAGTCTCAAGATCCAGCGCTTCAAG---CCCACCG 2137  
Qy 795 sIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspG1 815  
Db 2138 CATCATCTTCTACCGCGAGCGTCTCTGAAGCCAGTTCAGCAGGTCTCCACACGA 2197  
Qy 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835  
Db 2198 GTTGCTGGCCATCGTGAGCGCTGTATCAAGCTA-----GAAAAAGACTA 2242  
Qy 835 oGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgAr 855  
Db 2243 CCAGCCCGGATCACCTTCACTGTTGTCAGAGAGGACCCACCGCGCTCTTC---TG 2299  
Qy 855 gMetGluLysAspLysProValValaLysAspLeuThrProAlaGluThrAspValAl 875  
Db 2300 CACTGACAGACGAGCGGTGGGAAAGTGGAACATTCACGA----- 2345  
Qy 875 aValAlaAlaValLysGlnTrpGluLysAspMetLysGluSerLysGluThrGlyIleVa 895  
Db 2345 ----- 2345  
Qy 895 lAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPh 915  
Db 2346 -----GGCAGCACTGTGACACAGAAATCAACCCACCCAGCGGTTCGACT 2392  
Qy 915 ePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935  
Db 2393 CTACCTGTGTAGTCACGCTGGCATCCAGGAGACAGAGCGCTTCGACTATCACGTCCT 2452  
Qy 935 tTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPh 955  
Db 2453 CTGGGACGCAATCGTTTCTCTCTGATGAGTCGACATCTCAACCTACAGCTGTGTCA 2512  
Qy 955 eLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHisLe 975

Db 2513 CACCTACGTGGCTGCACACGCTCCGTGTCCATCCCGCCAGCCAGCATACTACGCTCACCT 2572  
Qy 975 uSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990  
Db 2573 GGTGGCCTTCGGGCCAGG---TACCACCTGGTGGATGAAGNACAT 2615  
RESULT 13  
ABS78717  
ID ABS78717 standard; cDNA; 3580 BP.  
XX  
AC ABS78717;  
XX  
DT 16-DEC-2002 (first entry)  
XX  
DE Human cDNA encoding NAAP7, from INCYTE no.1725129CB1.  
XX  
KW Human; ss; gene; nucleic acid associated protein; NAAP; cancer;  
KW cell proliferative disease; cancer; atherosclerosis; hepatitis;  
KW neurological disorder; Parkinson's disease; Alzheimer's disease; stroke;  
KW epilepsy; developmental disorder; renal tubular acidosis; anaemia;  
KW glaucoma; hypothyroidism; autoimmune disorder; AIDS;  
KW inflammatory disorder; acquired immunodeficiency syndrome; allergy;  
KW atopic dermatitis; arthritis; bacterial infection; viral infection;  
KW parasitic infection; protozoal infection; fungal infection.  
XX  
OS Homo sapiens.  
XX  
XX WC200272630-A2.  
XX  
PD 19-SEP-2002.  
XX  
XX 07-FEB-2002; 2002WO-US003844.  
XX  
XX 09-FEB-2001; 2001US-0268118P.  
PR 21-FEB-2001; 2001US-0270963P.  
PR 22-FEB-2001; 2001US-0270858P.  
PR 23-FEB-2001; 2001US-0271194P.  
PR 07-MAR-2001; 2001US-0274071P.  
PR 12-APR-2001; 2001US-0283496P.  
PR 09-NOV-2001; 2001US-0344650P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Thornton M, Hafalia AJA, Lu DAM, Arvizu C, Swarnakar A, Lu Y;  
PI Warren BA, Baughn MR, Tang YT, Lee EA, Yao MG, Ramkumar J, Khan FA;  
PI Gandhi AR, Ding L, Yue H, Gietzen KJ, Walia NK, Thangavelu K;  
PI Elliott VS, Marquis JP;  
XX  
XX WPI; 2002-723320/78.  
DR P-PSDB; ABG97473.  
XX  
XX New human nucleic acid-associated proteins (NAAP), useful for diagnosing,  
PT treating and preventing diseases or conditions associated with the  
PT aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,  
PT infections.  
XX  
XX Claim 5; Page 158-159; 162pp; English.  
XX  
XX The invention relates to an isolated polypeptide comprising one of 10  
CC human nucleic acid associated protein (NAAP1-10), or a biologically  
CC active or immunogenic fragment of the polypeptide, and their encoding  
CC nucleic acid. Also included are a recombinant polynucleotide comprising a  
CC promoter sequence operably linked to the polynucleotide, a cell  
CC transformed with the recombinant polynucleotide, a transgenic organism  
CC comprising the recombinant polynucleotide, an anti-NAAP antibody,  
CC screening for a compound that is effective as an antagonist or modulator  
CC of NAAP, generating an expression profile of a sample containing the  
CC polynucleotides and an array comprising different nucleotide molecules  
CC affixed on a solid substrate, nucleotide molecule comprises a first  
CC oligonucleotide or polynucleotide sequence specifically hybridisable with  
CC at least 30 contiguous nucleotides of the target (NAAP) polynucleotide.  
CC The polypeptides and polynucleotides are useful in diagnosing, treating

CC and preventing diseases or conditions associated with the decreased  
 CC expression or overexpression of NAAP, such as cell proliferative diseases  
 CC (e.g. cancer, atherosclerosis, hepatitis), neurological disorders  
 CC (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),  
 CC developmental disorders (renal tubular acidosis, anaemia, glaucoma,  
 CC hypothyroidism), autoimmune/inflammatory disorders (AIDS (acquired  
 CC immunodeficiency syndrome), allergies, atopic dermatitis, arthritis) and  
 CC infections (e.g. bacterial, viral, parasitic, protozoal, fungal) and many  
 CC other diseases and disorders listed in the specification. These are also  
 CC useful in assessing the effects of exogenous compounds on the expression  
 CC of nucleic acid and amino acid sequences of NAAP. The NAAP or its  
 CC fragments are useful in screening compounds for effectiveness as agonist  
 CC or antagonist of the polypeptides, or in altering the expression of the  
 CC target polynucleotide and compounds that specifically bind to or modulate  
 CC the activity of the polypeptide. The microarray is useful in monitoring  
 CC or measuring protein-protein interactions, drug-target interactions, and  
 CC gene expression profiles. The present sequence encodes an NAAP protein  
 XX

SQ Sequence 3580 BP; 903 A; 953 C; 926 G; 797 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 2,34e-42 Length: 3580  
 Score: 566.00 Matches: 239  
 Percent Similarity: 40.0% Conservative: 159  
 Best Local Similarity: 24.0% Mismatches: 385  
 Query Match: 10.6% Indels: 214  
 DB: 6 Gaps: 38

US-10-645-746-3 (1-1020) x ABS78717 (1-3580)

QY 19 ProGluMetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLys 38  
 DB 95 CCATCCCAAGGATATGCTTCAAGCTCCACCTAGACCCGATCTGGGACCTCCGGGAGA 154  
 QY 39 LysValLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyr 58  
 DB 155 ACAATCAAAATACAGGCCAATTTCTTGAATGGACATCCCAAAATTCGACATCATCAT 214  
 QY 59 TyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPhe 78  
 DB 215 TATGAATGGATATCAAGCCAGAGAG-----TGC 244  
 QY 79 ProLysLysThrGluIleProLysProAspArgAlaLysLeuPheTrpGlnHisLeuArg 98  
 DB 245 CCAGAGAGATTAAACAGGGAATCTGGGAACATG-----GTCCAG 286  
 QY 99 HisGluLysLysGlnThrAspPheLysLeuGluAspTyr-----ValPheAspGluLys 116  
 DB 287 CACTTTAAACACAG-----ATCTTTGGGATCGGAAGCCGCTGTTGACGCGCAGG 337  
 QY 117 AspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGlu 136  
 DB 338 AAGAATCTATACAGCCATGCCCTTCGANTGGAGGCGCAAGGTG-----GAG 388  
 QY 137 LysValValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeu 156  
 DB 389 CTGGAGGTACCGCTCCAGAGAGAGCAAGGATCG-----ATCTTCAAGGTGTCATC 442  
 QY 157 TyrThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnPro 176  
 DB 443 AAGTGGGTCTCTCGTGAGCTTGACGGGTTACAGCATGTCATTTTCAGGGCGGTGCC 502  
 QY 177 GluLysAspGluLysAlaAsnArgSerTyrLysPheLysLysAsnValMetThrGln--- 195  
 DB 503 AGCGTCCCTTTTCAG-----ACGATCCAGGCCCTGGACGTGTCATGAGGCACCTTG 553  
 QY 196 ---LysValArgTyrAlaProPheValAsnGluLysValGlnPheAlaLysAsn 214  
 DB 554 CCATCCATGAGGTACACCCCGTG-----577  
 QY 215 PheValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArg 234  
 DB 578 -----GGCGGCTCTTCTTCCAGCGGTCCGAAGGCTGCTCTAACCCCT----- 619

QY 235 PheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGly 254  
 DB 620 -----CTTGGCGGGCGGCGAAGGTGTGTGTGGCTTCCATCAGTCC 661  
 QY 255 IleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPhe 274  
 DB 662 GTCCGGCTCTCTCTC---TGGNAATGATGCTGATATTTGATGTGTGCAGCAACAGCGTTT 718  
 QY 275 TyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuLeuLeuValAspProGlnSer 294  
 DB 719 TACAAGGCA-----CAGCCAGTAACTCGAGTTTGTGTGAAGTTTTCGATTTTAAAGT 772  
 QY 295 CysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIle 314  
 DB 773 ATTTGAAGACAA-----CRAAAACCTCTG 796  
 QY 315 ArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLysLys----- 332  
 DB 797 ACAGATTCCCAAGGGTAAAGTTTACCAAGAAATTAAGGTCTTAAGGTGGAGATAACG 856  
 QY 333 ---CysAlaGluValTrpAspAsn-----GluMetSerArgLeuThrGluArg 347  
 DB 857 CACTGTGGCAGATGAAGAGGAAGTACCGCGTCTGCAATGTGACCGCGCGCGCCAGT 916  
 QY 348 HisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLys 367  
 DB 917 CACCAACATTC---CCGCTGACGAGGAGCGGCGGACCGTGGAGTGACAG----- 967  
 QY 368 SerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGluGlu 387  
 DB 968 -----GTGSCCCAGTAT-----TTCAAG-----GACAGG 991  
 QY 388 AsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLys 407  
 DB 992 CACAAGTTGGTTCTCGCTACCCCACTCCCATGTTTACAAGTCGGACAGGACGAGAA 1051  
 QY 408 GluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsn 427  
 DB 1052 CACACTTACCTTCCCTCGAGGTCTGTAACTTTGGCAGGCAAGATGATTATTAANA 1111  
 QY 428 ArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr---ArgLysProHis 446  
 DB 1112 TTAACGGAC-----AATCAGACCTCAACCATGATCAGACCGACTGTAGTGGCGGCC 1165  
 QY 447 AspTyrLysGluAsnThrLeuLysMetLysGluLeuAspPheSerSerSerGluGluLeu 466  
 DB 1166 GATCGCAAGAAGAGATTAGCAAAATTGATCGAAGTGCAGATTTCAACACAGATCCA--- 1222  
 QY 467 AsnPheValGluArgPheGlyLysCysSerLysLeuGlnMetIleGluCysProGlyLys 486  
 DB 1223 ---TACGTCCGTGAATTTGGATCATGTGTCAAGATGAGATGACAGACGTGACTGGCGG 1279  
 QY 487 ValLysLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrProVal 506  
 DB 1280 GTGCTGACGCGCCCTCCATCTCTACGGGGGCGAGAAATAAGATTTCGACACCCCTGTC 1339  
 QY 507 -----IleArgGlyPheGlnGluLysGlnLeuAsnVal---ValProLysLysGluLeu 523  
 DB 1340 CAGGCGCTCTGGGACATCGGAAACAGCAGTTTCCACCGGCATCGAGATCAAGGTGTGG 1399  
 QY 524 CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp 543  
 DB 1400 GCCATTGGTGTCTCGCCCCCAG-----CGCCAGTGCACGGAAGTCCAT--- 1444  
 QY 544 ValValLysPheTyrThrGluLeuIle-----GlyGly-CysLysPhe 557  
 DB 1445 ---CTGAAGTCTTTCACAGAGCAGCTCAGAAAGATCTCGAGAGACGCTGGCATGCCATC 1501  
 QY 557 eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIle-----MetTy 575  
 DB 1502 CAGGCGCCAGCGGTCTTCTGCAAAATACGCG-CAGGGGGCGGACGCGTGGAGCCCATGTT 1560



575 rAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleG1 595  
1561 CCGCAGCTGAGAACACGATGCG----- 1585  
595 yArgPheGluIleAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysG1 615  
1585 ----- 1585  
615 uGlnLysValLeuMetPheIleIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheVa 635  
1586 ---GGCCTGCAGCTGGTGGTGGTCTCATCTGCCCGCAAGACGCGCTAGCGGAGGT 1641  
635 lLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrVa 655  
1642 CAGCGCGTGGGAGACACGCTGCTGGGGATGGCCAGCGAGTGGTGCAGATGAAGAAGCT 1701  
655 lThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnI1 675  
1702 G-----CAGAGGACCCAGCCACAGACCCCTGTCCAACCT 1734  
675 eAlaLeuLysIleAsnAlaLysLeuGlyIleAlaAsnGlnGluLeuAspTrpSerGluI1 695  
1735 CTGCCTGAAGATCAACGCTCAAGCTGGGAGCGGTGAACAACATCCTG----- 1780  
695 eAlaGluIleSerProGluGluLysGluArgLysThrMetProLeuThrMetTyrVa 715  
1781 -----CTGCCCCAGGCGGAGCGCGCGGTTCACAGCGCCGTC---ATCTTCT 1827  
715 lGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaLaVa 735  
1828 GGGAGCAGACGTCCTACCCCGCGCGGGATGGGAAGACCCCTCCATTGCCGCGT 1887  
735 lValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluG1 755  
1888 GGTGGGAGCATGAGACGCCCACTGCTACTGCGCCACCGTGGCGTGCAGCAG-- 1945  
755 uCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAl 775  
1946 -----CACGCGCAGGAGATCATCAAGACCTGGCCGC 1977  
775 aLysPheValLysLeuArgGluPheAlaGluAsnAspAsnArgAlaProAlaH1 795  
1978 CATGGTCCGGAGCTCTCTCATCCAGTCTTACAGTCCACGGCTTCAAG---CCACCCG 2034  
795 sIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspG1 815  
2035 CATCATCTTCTACCGGCGAGCGTCTCTGGAAGCCAGTTCCAGCAGGTTCTCCACCACGA 2094  
815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835  
2095 GTTGTGGCCCTCCGTCGAGCGCTGTATCAAGCTA-----GAAAGAGCTA 2139  
835 oGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgAr 855  
2140 CCAGCCGGGATCACCTTCTATCTGTGTGAGAGAGGACACACCGGCTTCTC---TG 2196  
855 gMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAl 875  
2197 CACTGACAGACGAGCGGTTGGGAAAGTGGAAACATTCACGCA----- 2242  
875 aValAlaAlaValLysGlnTrpGluLysMetLysGluSerLysGluThrGlyIleVa 895  
2242 ----- 2242  
895 lAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPh 915  
2243 -----GGCAGCACTGTGGACAGAAATCACCACCCACCGAGTTTCACCTT 2289  
915 ePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935  
2290 CTACCTGTGTAGTCAGCTGGCATCCAGCGGACAGAGCGCTTCGCACTATCAGCTCT 2349  
935 tTyrAspAspLysGlyMetSerClnAspGluValTyrLysMetThrTyrGlyLeuAlaPh 955

2350 CTGGGAGCAATCGTTTCTCTCTGATGAGCTGCAGATCCTAAACCTACCAGCTGTGTCA 2409  
955 eLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHisLe 975  
2410 CACTAGCTGGCTGCACAGCTCCGTGTCATCCAGCGGCAGCATACTACGCTCACT 2469  
975 uSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990  
2470 GGTGGCTTCGGGCCAGG---TACCACCTGGTGGATAAGGAACAT 2512  
RESULT 14  
AAH14510  
ID AAH14510 standard; cDNA; 3996 BP.  
XX  
AC AAH14510;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:12038.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-00116126.  
XX  
XX 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isoqai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;  
PI Iahii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
PS  
PS Claim 8; SEQ ID NO 12038; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention





```

Db      1745 GTCCAACTCTGCTCAAGATCAATGTCAAACTGGTGGCATTAACAGCATCTTA----- 1799
QY      692 pSerGluLeuAlaGluLeuSerProGluGluGluGluArgArgLysThrMetProLeuTh 712
Db      1800 -----GTCCCAACACAGCGCTCTGCCGTTTTCACACAGCGCAGT-- 1838
QY      712 rMetTyrValGlyLeuAspValThrHisProThrSerTyrSerGlyLeuAspTyrSerIl 732
Db      1839 -ATATTCCTGGGAGCAGATGTACACACCCCCCAGCAGGGGATGGGAAAAAACCTTCTAT 1897
QY      732 eAlaAlaValAlaAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValTh 752
Db      1898 CACAGCAGTGTAGGCGATGTAGTATGCCACCCAGCCGATATCTGTCTACTGTGCGGT 1957
QY      752 rGlnGluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIl 772
Db      1958 ACAG-----CGACCAGCGCAAGAGAT 1978
QY      772 eLeuGlu-----AlaLysPheValLys---LeuLeuArgGluPheAlaGluAsnAs 789
Db      1979 CATTGAAGACTTGTCTACATGTGTGGTGGCTCTCTATCCCAATTTCTACAAGTCCACCCG 2038
QY      789 pAsnArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLe 809
Db      2039 TTTCAG---CCTACCGCATATCTCTCCGAGATGGGTGCTGAGAGCCAGCTACC 2095
QY      809 uArgValSerHisAspGluLeuArgSerLeuLysSerGluValGlnPheMetSerGl 829
Db      2096 CCAGATACTCCACTATGAGCTACTGGCCATTCTGTGATCGCTGCATCAAACTG----- 2147
QY      829 uArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHisAs 849
Db      2148 -----GAAAGGACTACACAGCTGGATCATCTATATTGTGTGGTGCAGAAACGCCATCA 2200
QY      849 nThrArgLeuLeuArgMetGluLysAspLysProValValAsnLysAspLeuThrPr 869
Db      2201 CACCCGCTTTTC---TGCTGCTACAGAGATGGCGAATGGGAGAGTGTGAACATCCC 2257
QY      869 oAlaGluThrAspValAlaValAlaValLysGlnThrGluGluAspMetLysGluSe 889
Db      2258 AGCT----- 2261
QY      889 rLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeuIleValSe 909
Db      2262 -----GGGACCACAGTGGGACACCAACATCACCCA 2290
QY      909 rLysTyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGlyThrSerArgPr 929
Db      2291 CCATTTGATTGATTTCTATCTGTGCACCCAGCGGATCCAGGGCACCAGCCGACC 2350
QY      929 oGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMe 949
Db      2351 ATCCCATTAATACTTTTGGGATGACAACCGTTTCACAGCAGATGAGCTCCAGATCCT 2410
QY      949 tThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValPr 969
Db      2411 GAGGTACCACTGTGCACACTTACGTACGATGCACACGCTCTGTCTATCCAGCACC 2470
QY      969 oValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGl 989
Db      2471 TGCTTACTATGCCGCTGTGTGCTTCCGGGACCGA---TACCACCTGTGTGACAGGA 2527
QY      989 uHisTyrIleGlyAsp 994
Db      2528 GCATGACAGTGGAGAG 2543

```

RESULT 15

ABSS1392

ID ABSS1392 standard; cDNA; 2955 BP.

XX AC

XX ABSS1392;

XX

```

DT      05-NOV-2002 (first entry)
XX      cDNA encoding larval viability associated protein #15.
XX      Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;
XX      rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar beet;
XX      oilseed rape; soybean; vegetable crop; fruit; gene; 88.
XX      Drosophila melanogaster.
XX      WO200257455-A2.
XX      25-JUL-2002.
XX      18-JAN-2002; 2002WO-US001568.
XX      18-JAN-2001; 2001US-0262351P.
XX      (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX      Stam L, Bachmann J, Broadus J, Kamdar KP;
XX      WPI; 2002-590746/63.
XX      P-PSDB; ABG70016.
XX      Identifying inhibitors of activity of proteins essential for Drosophila
XX      larval viability comprises expressing in a host a protein essential for
XX      larval activity and identifying compounds that inhibit or interact with
XX      the protein.
XX      Claim 1; Page 91-93; 169pp; English.
XX      The invention describes a method of identifying compounds that inhibit
XX      the activity of, or that interact with a protein essential for Drosophila
XX      larval viability comprising expressing in a recombinant host a DNA
XX      molecule to produce a protein essential for larval viability. The method
XX      is useful for identifying compounds with insecticidal activity. Compounds
XX      identified are useful as insecticides in crops such as maize, wheat,
XX      oats, rye, sorghum, rice, barley, millet, turf, cotton, sugarcane, sugar
XX      beet, oilseed rape, soybeans, vegetable crops and fruits. This sequence
XX      encodes a fruit fly larval viability associated protein
XX      Sequence 2955 BP; 651 A; 901 C; 807 G; 596 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      2,48e-42      Length:      2955
Score:          564.50      Matches:      233
Percent Similarity: 40.4%      Conservative: 174
Best Local Similarity: 23.1%      Mismatches: 394
Query Match:    10.6%      Indels:      206
DB:             6          Gaps:      37

US-10-645-746-3 (1-1020) x ABSS1392 (1-2955)
QY      38 LysLysValLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGlu 57
Db      436 CGCCCGATTGTGTGCGCCCAATCATCTCCAGGTGACA-----ATGCCCGGTGGC 486
QY      58 Tyr---TyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLys 76
Db      487 TATGTGCATCATATGACATCATATACAGCCGACAGTGTCCGGAAGGTGACCGT 546
QY      77 ProPheProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHis 96
Db      547 GAGATT-----ATCGAGACTATGGTGCATGCTATAGCAAGATCTTCGGA----- 591
QY      97 LeuArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLys 116
Db      592 -----GTGCTCAAGCCG---GTGTTTCGATGTGTCG 618
QY      117 AspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGlu 136
Db      619 AACATCTGTATACCCCGCATCCCTGCCATGGCAACGAGCGTCTA-----GAG 669

```

```
QY 137 LysValValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeu 156
DB 670 CTGGAGGTTACTCTACCCCGCGCGGCAAGATCGA----- 705
QY 157 TyrThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnPro 176
DB 706 -----ATCTTCGCGTGACGATCAAGTGGCAGGCTCAGGTCTCGCTCTCAATTGG 756
QY 177 GluLysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLys 196
DB 757 GAGGAAGCTCTCGAAGCGCGC-----ACGCGGCAG 786
QY 197 ValArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheVal 216
DB 787 ATACCCATGATGCCATTTTGGCGTCGATGTGGTCATCGGCCATCGCCAGCAGTACG 846
QY 217 TyrAsp-----AsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArg 234
DB 847 TACACGCCAGTGGAGCTAGCTTCTTCAGTTCCTCCGAGGGTTACTACCATCCC----- 900
QY 235 PheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGly 254
DB 901 -----CTGGGTGTGCAGCGAGGTTTGGTTCGGTTTCCATCAGAGC 942
QY 255 IleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPhe 274
DB 943 GTAAGGCC-----TCGCAGTGGAGAGATGCTCAATATCGATGTCTGGCGCCACCGCTTTC 999
QY 275 TyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuIleValAspProGlnSer 294
DB 1000 TACAAGGCT-----CAACCACTGATTCATTCATGTGCGAGGTGCTGCACATTCGCGAC 1053
QY 295 CysAsnAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIle 314
DB 1054 ATCAACGAGCAG---CGCAACCGCTC----- 1077
QY 315 ArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLysCysAla 334
DB 1078 ACCGATTCGCAGCGCGTCAAGTTACGAGGAGATCAAGGGTTTGAAGATCAGATCACC 1137
QY 335 GluValTrpAspAsnGluMetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeu 354
DB 1138 CAC----- 1140
QY 355 CysGluGluAsnSerLeuValTyrLysValThrGlyLysSerAspArgly----- 371
DB 1141 TCGCGCCAGATCGCTCGCAAGTATCGTGTGTGCAACGTCACCTCGCGCGCCGCTCAGATG 1200
QY 372 -----ArgAsnAlaLysLysTyrAspThrThrLeuPheLysIle 384
DB 1201 CAATCAATCCCACTGAGTGGAGAACGACGACGCGTAGATGACCGTGGCCCAAGTAC 1260
QY 385 Tyr---GluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLys 403
DB 1261 TTCCTGGACAAGTACCGCATGAAATTCGCTACCGCATCTGCCCTGCTGCGAGTGGC 1320
QY 404 SerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGln 423
DB 1321 CAAGAGCACACACACTTACCTGCTCTAGAGGTGTGCAACATT---GTGCGCGGACAG 1377
QY 424 ArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArg 443
DB 1378 CGGTGCATTTAAAGCTGACCATATGACAGCGTGCACATCATCAAGGCCACAGCTCGT 1437
QY 444 LysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSer 463
DB 1438 TCAGCTCCGGATCGTGAGCGTGAGATTAACAATCTGGTAAAGCGCGCCGACTTCAACAAC 1497
QY 464 GluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCys 483
DB 1498 GAT-----TCGTATGTGCAAGAGTTTGGCCTGACCATCTCCAATTCGATGATGAGGTA 1551
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 6, 2006, 01:16:25 ; Search time 2698 Seconds

(without alignments)  
6968.155 Million cell updates/sec

Title: US-10-645-746-3

Perfect score: 5349

Sequence: 1 MSSNFPLEKGFYRHSIDPE.....RHEMHPLQTNVKGMSFA 1020

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/abs/ABSSWEB.spool/US10645746/runat\_05072006\_110158\_5297/app\_query.fasta\_1  
-DB=Published Applications.NA.Main -OPMT=fastap -SUFFIX=p2n.rnpbm  
-MINMATCH=0.1 -LOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext  
-HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs804  
-USER=US10645746 @CGN 1.1 1780 @runat\_05072006\_110158\_5297 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications.NA.Main.\*

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq.\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq.\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5349	100.0	3227	9	US-10-645-746-2

2	5349	100.0	3227	10	US-10-645-735-2	Sequence 2, Appli
3	5349	100.0	3227	16	US-11-144-985-2	Sequence 2, Appli
4	5001	93.5	3719	9	US-10-645-746-1	Sequence 1, Appli
5	5001	93.5	3719	10	US-10-645-735-1	Sequence 1, Appli
6	5001	93.5	3719	16	US-11-144-985-1	Sequence 1, Appli
7	577	10.8	7478	6	US-10-007-078-3	Sequence 3, Appli
8	577	10.8	7478	13	US-11-035-669-3	Sequence 3, Appli
9	575.5	10.8	2910	3	US-09-533-029-105	Sequence 105, App
10	575.5	10.8	2910	6	US-10-286-264-101	Sequence 101, App
11	575.5	10.8	2910	8	US-10-374-780A-2083	Sequence 2083, Ap
12	575.5	10.8	2910	8	US-10-412-699B-751	Sequence 751, App
13	566	10.6	3011	8	US-10-115-635-50	Sequence 50, Appl
14	566	10.6	3580	8	US-10-467-397-17	Sequence 17, Appl
15	564.5	10.6	3486	13	US-11-097-143-27926	Sequence 27926, A
16	564.5	10.6	3586	13	US-11-097-143-27929	Sequence 27929, A
17	564.5	10.6	3743	13	US-11-097-143-14510	Sequence 14510, A
18	556.5	10.4	2914	3	US-09-774-434-2	Sequence 2, Appli
19	556.5	10.4	3050	7	US-10-175-492-4	Sequence 4, Appli
20	556.5	10.4	3578	10	US-10-756-149-37	Sequence 37, Appl
21	550.5	10.3	3307	9	US-10-852-630A-4	Sequence 4, Appli
22	550.5	10.3	3325	10	US-10-483-505-66	Sequence 66, Appl
23	536.5	10.0	3509	8	US-10-437-963-89857	Sequence 89857, A
24	533.5	10.0	3509	7	US-10-437-963-18974	Sequence 18974, A
25	530.5	9.9	2815	7	US-10-269-909-20	Sequence 20, Appl
26	519.5	9.7	8465	13	US-11-097-143-27928	Sequence 27928, A
27	519.5	9.7	10149	13	US-11-097-143-27925	Sequence 27925, A
28	519.5	9.7	12558	13	US-11-097-143-14509	Sequence 14509, A
29	519	9.7	3147	8	US-10-374-780A-2085	Sequence 2085, Ap
30	509.5	9.5	2973	8	US-10-437-963-6231	Sequence 6231, Ap
31	507	9.5	3310	8	US-10-437-963-21837	Sequence 21837, A
32	507	9.5	3372	8	US-10-437-963-6229	Sequence 6229, Ap
33	502	9.4	3474	8	US-10-437-963-98130	Sequence 98130, A
34	501	9.4	2967	3	US-09-938-842A-432	Sequence 432, App
35	501	9.4	2967	3	US-09-938-842A-432	Sequence 432, App
36	499.5	9.3	2832	8	US-10-437-963-35635	Sequence 35635, A
37	499.5	9.3	3761	8	US-10-437-963-55972	Sequence 55972, A
38	496.5	9.3	3718	8	US-10-425-114-20232	Sequence 20232, A
39	495	9.3	2042	6	US-10-174-363-1	Sequence 1, Appli
40	495	9.3	2042	13	US-11-093-888-1	Sequence 100929,
41	494.5	9.2	3994	6	US-10-437-963-100929	Sequence 53, Appl
42	489	9.1	3705	13	US-11-093-888-53	Sequence 53, Appl
43	489	9.1	3705	13	US-11-093-888-53	Sequence 53, Appl
44	488	9.1	3323	7	US-10-225-066A-151	Sequence 151, App
45	488	9.1	3323	8	US-10-374-780A-237	Sequence 237, App

#### ALIGNMENTS

RESULT 1  
US-10-645-746-2  
; Sequence 2, Application US/10645746  
; Publication No. US20040265839A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE  
; TITLE OF INVENTION: INTERFERENCE  
; FILE REFERENCE: UMY-052DV1  
; CURRENT APPLICATION NUMBER: US/10/645,746  
; CURRENT FILING DATE: 2003-08-20  
; PRIOR APPLICATION NUMBER: US 09/689,992  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/193,218  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: US 60/159,776  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 3227  
; TYPE: DNA

; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (21) ... (3080)  
; US-10-645-746-2

## Alignment Scores:

Pred. No.: 0 Length: 3227  
Score: 5349.00 Matches: 1020  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 9 Gaps: 0

US-10-645-746-3 (1-1020) x US-10-645-746-2 (1-3227)

QY 1 MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspProGlu 20  
DB 21 ATGCTCTCGAATTTTCCCGAATTGGAAAGGATTTTATCGTCATTTCTCGATCCGGAG 80  
QY 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysVal 40  
DB 81 ATGAAATGGCTTTCGAGGCCCACTGGTAAATGCGCGCAAAATTTCTATGAGAAGAAGTA 140  
QY 41 LeuLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrGlu 60  
DB 141 CTTCTTTTGGTAAATTTGGTTCAAGTTCTCAGCAAAATTTTATCGATCGGGAATACTACGAG 200  
QY 61 TyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLys 80  
DB 201 TATGAAGTGAATAATGACAAAGGAAGTATTGAAATAGAAAACCGAGAAAACCTTTCCCAAAA 260  
QY 81 LysThrGluLysProLysProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGlu 100  
DB 261 AAGACAGAAATTTCCCAATTTCCGATCGTCAAAATCTTCTCGCAACATCTTCGGCATGAG 320  
QY 101 LysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAspThrValTyr 120  
DB 321 AAGACGACACAGATTTTATTTCTCGAAGACTATGTTTTGTGTAAGAAAGGACACTGTTTAT 380  
QY 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysValValLys 140  
DB 381 AGTGTGTGTGCGTGAACACTGTGCATCAAAAATGCTGGTTTCGGAGAAAGTAGTAAAA 440  
QY 141 LysAspSerGluLysLysAspGluLysAspLeuGluLysIleLeuTyrThrMetIle 160  
DB 441 AAGGATTCGGAGAAAAGATGAAAGGATTTGGAGAAAAAATCTTATACAAATGATA 500  
QY 161 LeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGluLysAspGlu 180  
DB 501 CTTACCTATCGTAAAAAATTTCACTGAACTTTAGTCGAGAAAATCCGGAAAAAGACGAA 560  
QY 181 GluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysValArgTyrAla 200  
DB 561 GAAGCGAATCGGAGTTTACAAATTCCTGAAGATTTTATGACCCAGAAAGTTTCGTCACGG 620  
QY 201 ProPheValAsnGluLysValGlnPheAlaLysAsnPheValTyrAspAsn 220  
DB 621 CTTTGTGAAACGAGAGATTTAAAGTACAAATTCGCGAAAAATTTTGTGTACCAATATAAT 680  
QY 221 SerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGlnSerLeuGlu 240  
DB 681 TCATTTCTCGAGTTCTCGATGTTTTCAGATCCAAACAGATTCGAACATCATTAGAA 740  
QY 241 ValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyLysGluLeuPheAsp 260  
DB 741 GTAGCACCAGAAATCGAAGCTGTTTTCAGATCCAAACAGATTCGAACATCATTAGAA 800  
QY 261 GlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAlaProLysMet 280  
DB 801 GTGGAACCTGTGCTCAATTTTGCATTTGTGCAATAACTATTCTACAACTACCGAAAAATG 860  
QY 281 SerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSerCysAsnAspValArg 300

DB 861 TCTCTTCTGGATTATCTTCTCTCTAATTTGTCGACCCCGCTGCTGTAACGATGATGTACGA 920  
QY 301 LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgPro 320  
DB 921 AAAGATCTTAAAAACAACACTGATGCGGGAATAATGACAATCAGACAACGCGCGGCCCA 980  
QY 321 ArgIleArgGlnLeuLeuGluAsnLeuLysCysAlaGluValTrpAspAsnGlu 340  
DB 981 AGAATTCGACAAATTTTGGAAATTTGAAGCTGGAATGCGCAAGTTTGGGATATACGAA 1040  
QY 341 MetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeuLysGluGluAsnSerLeu 360  
DB 1041 ATGTCGAGATTGACAGAACGACATCTGACATTTCTAGATTTTGTGCGAGGAAAACTCTCTT 1100  
QY 361 ValTyrLysValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThr 380  
DB 1101 GTTTATAAGTCACTGGTAAATCGGACAGAGAGAAATGCAAAAAGTAGTACGTACTACA 1160  
QY 381 LeuPheLysIleTyrGluGluAsnLysPheIleGluPheProHisLeuProLeuVal 400  
DB 1161 TTGTTCAAAATCTATGAGAAACAAAGATTCATTTGAGTTTCCCACTACCTACCTAGTC 1220  
QY 401 LysValLysSerGlyValLysGluTyrAlaValProMetGluHisLeuGluValHisGlu 420  
DB 1221 AAAGTTTAAAGTGGAGCAAAAGATACGCTGTACCAATGGAACATCTTCAAGTTTCATGAG 1280  
QY 421 LysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLysArg 440  
DB 1281 AAGCACAAAGATACAGAATCGAATTCATCTGGTGATGCAAGCAAGTTTCTTAAAGCGCA 1340  
QY 441 AlaThrArgLysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAsp 460  
DB 1341 GCTACACGAAACCTTCAGACTACAAAGAAATACCTTAAATGCTGAAAGAAATTTGGAT 1400  
QY 461 PheSerSerGluGluLeuAsnPheValGluArgPheGlyLeuLysSerLysLeuGlnMet 480  
DB 1401 TTTCTCTCTGAAGAGCTAAATTTTCTGAAAGATTTGGATTTATGCTCCAAATTCAGATG 1460  
QY 481 IleGluCysProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGln 500  
DB 1461 ATCGAATGTCCAGGAAAGGTTTGAAGAGCCCAATGCTTGTGAATAGTGTGAATGAACAA 1520  
QY 501 IleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGlu 520  
DB 1521 ATTAATAATGACACCACTGATTCGTGATTTCAAGAAACAAATTTGAATGCTGTTCCGAA 1580  
QY 521 LysGluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGlu 540  
DB 1581 AAAGAACTTTGCTGTCTGTTTTGTAGTCAACGAAACAGCGGGAATCCATGCTTAGAA 1640  
QY 541 GluAsnAspValValLysPheTyrThrGluLeuIleGlyCysLysPheArgGlyIle 560  
DB 1641 GAGAACGACGTTGTTAAGTTCTACACCGAACTTAATTTGGTGGTTGCAAGTTCCGTGAATA 1700  
QY 561 ArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAsn 580  
DB 1701 CGAATTTGGTCCCAATGAAAAACAGAGAGCGCAATCTATTATGTCAGCCCGACGAAAAAT 1760  
QY 581 GluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAla 600  
DB 1761 GAATATGCTTCTACAAAAATTTGTACACTAAATACCGAATCGTAGATTTGGAATAGCC 1820  
QY 601 AlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGlnLysValLeuMet 620  
DB 1821 GCACAGAGCGGAAGATATGTTGAACGCTTCTCCCGATAAAGAAACAAAGTCTTTAATG 1880  
QY 621 PheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAsp 640  
DB 1881 TTCAATATCATTTCCAAACGACAACTGAATGCTTACGGTTTTGTGAAACATTTATTCGAT 1940  
QY 641 HisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAla 660

Db 1941 CACACCATCGGTGTAGCTAATCAGCATATTAATCTTCTGAACACAGTCACAAAAGCTTTGGCA 2000  
Qy 661 SerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsn 680  
Db 2001 TCACTAAGGCACGAGAAAGGATCAAAACGAATTTTCTATCAATCAAAATTCGAAATCAAC 2060  
Qy 681 AlalysLeuGlyGlyIleAsnGlnGluLeuAspTrpSerGluIleAlaGluIleSerPro 700  
Db 2061 GCGAAATTAGAGGTATTAAACGAGGCTTGACTGGTCAGAAATTCGAGAAATATCACCA 2120  
Qy 701 GluGluLysGluArgArgLysThrMetProLeuThrMetTyrValGlyIleAspValThr 720  
Db 2121 GAAGAAAAGAAAGACCGAAACCAATGCCATTAACTATGATGTGTGGAAATGATTAAT 2180  
Qy 721 HisProThrSerTyrSerGlyIleAspTyrSerIleAlaValValAlaSerIleAsn 740  
Db 2181 CATCCAACTCTCCTACAGTGGAAATGATTAATCTATAGCGGTGTAGTAGCGAGTATCAAT 2240  
Qy 741 ProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGluCysArgProGlyGlu 760  
Db 2241 CCAGGTGGAACTATCTATCGAAATATGATTTGTGACTCAAGAAGAAATGCTGTCGGGTGAG 2300  
Qy 761 ArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeu 780  
Db 2301 CGTGCAGTGGCTCATGGAACGGGAAGAAACAGATATTTTGGAGCAAGCTTCGTGAATTC 2360  
Qy 781 LeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAlaHisIleValValTyrArg 800  
Db 2361 CTCAGAGAAATTCGAGAAAACCAACGAAATCGAGCACCGCGCATATTTGTAGTCTATCGA 2420  
Qy 801 AspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLys 820  
Db 2421 GACGGAGTTAGCGAATTCGGAGATGCTACGTGTGTAGTCATGATGAGCTTCGATCTTAAAA 2480  
Qy 821 SerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLysTyrThr 840  
Db 2481 AGCGAGTAAACAAATTCATGTCGGAACGGATGGAGAAATCCAGAGCCGAAAGTACAGC 2540  
Qy 841 PheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgMetGluLysAspLys 860  
Db 2541 TTCATTGTGATTGAGAAAGACAAATACAGATTCGTTCTCGAAGATGGAAAAAGATAAG 2600  
Qy 861 ProValValAsnLysAspLeuThrProAlaGluThrAspValAlaValAlaValLys 880  
Db 2601 CCAGTGGTCAATAAAGATCTTATCTCCTGCTGAAACAGATGCTGCTGTGCTGTAAA 2660  
Qy 881 GlnTrpGluAspMetLysGluSerLysGluThrGlyIleValAsnProSerSerGly 900  
Db 2661 CANTGGAGGAGGATATGAAGAAGAAAGCAAGAACTGGAAATGTGAACCCATCATCCGA 2720  
Qy 901 ThrThrValAspLysLeuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHis 920  
Db 2721 ACAACTGTGATAAACTTATCGTTTCGAAATACAAAATTCGATTTTTCTTGGCATCTCAT 2780  
Qy 921 HisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGly 940  
Db 2781 CATGGTGTCTCTGGTACATCTCGTCCAGGACATTACATGTTATGTATGATGACGATAAGGA 2840  
Qy 941 MetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCys 960  
Db 2841 ATGAGCCAGATGAGTATTAATAAAGTACCTACGAGCTTCCTTTCTCTGCTAGATGT 2900  
Qy 961 ArgLysProIleSerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAla 980  
Db 2901 CGAAAAACCATCTCGTTGCTGTTTCGGTTTCATTATGCTCATTTATCATGTGAAAAAGCG 2960  
Qy 981 LysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThr 1000  
Db 2961 AAAGAGCTTTATCGAACTTACAAGGAACATTATCATCGGTGACTATGCACAGCCAGCACT 3020  
Qy 1001 ArgHisGluMetGluHisPheLeuGlnThrAsnValLysTyrProGlyMetSerPheAla 1020  
Db 3021 CGACACGAATGAGAACATTTCTCCAAACTTAACGTGAAGTACCCTGGAATGCTGTCGCA 3080

## RESULT 2

US-10-645-735-2  
; Sequence 2, Application US/10645735  
; Publication No. US20050100913A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; TITLE OF INVENTION: INTERFERENCE  
; FILE REFERENCE: UMG-052  
; CURRENT APPLICATION NUMBER: US/10/645,735  
; CURRENT FILING DATE: 2003-08-20  
; PRIOR APPLICATION NUMBER: US/09/689,992A  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/193,218  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: US 60/159,776  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 3227  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (21)...(3080)  
US-10-645-735-2  
Alignment Scores:  
Pred. No.: 0 Length: 3227  
Score: 5349.00 Matches: 1020  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 10 Gaps: 0  
US-10-645-746-3 (1-1020) x US-10-645-735-2 (1-3227)

Qy 1 MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspProGlu 20  
Db 21 ATGTCCTCGAAATTTCCCGAAATTTGGAAAAAGGATTTTATCGTCAATCTCTCGATCCGAG 80  
Qy 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysVal 40  
Db 81 ATGAAATGGCTTCGAGAGCCCACTGGTAAATGCGACGCAAAATCTATGAGAAAGATA 140  
Qy 41 LeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGlu 60  
Db 141 CTTCTTTTGGTAAATTTGGTTCAAGTTCTCCAGCAAAATTTACGATCGGGAATACTACGAG 200  
Qy 61 TyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLys 80  
Db 201 TATGAAGTGAATGCAAGGAAGTATTGAATAGAAAAACAGGAAAAACCTTTCCCAAAA 260  
Qy 81 LysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGlu 100  
Db 261 AAGACGAATTCCAATTCCTCGATCGTGGCAAACTCTTCTGGCAACATCTTCGGCATGAG 320  
Qy 101 LysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAspThrValTyr 120  
Db 321 AAGAAGCAGACAGATTTTATTTCTCGAAGACTATGTTTTTTCATGAAAGGACACTGTTTAT 380  
Qy 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysValLys 140  
Db 381 AGTGTGTGTCGACTGAACACTGTGCACATCAAAATGCTGCTTTTCGGGAAAGTAGTAAAA 440  
Qy 141 LysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeuTyrThrMetIle 160  
Db 441 AAGGATTCGAGAAAAAAGATGAAAGGATTTTGGAGAAAAAATCTTTATACAAATGATA 500



QY 161 LeuThrTyrArgLysPheHisLeuAenPheSerArgGluAenProGluLysAspGlu 180  
DB 501 CTTACCTATCGTAAATAATTTTACCTGAACTTTAGTCGAGAAATCCGAAAAAGACGAA 560  
QY 181 GluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysValArgTyrAla 200  
DB 561 GAAGCGAATCGGNGTTACAAATTCCTGAAGAATGTTATGACCCAGAAAGTTGCTACGGC 620  
QY 201 ProPheValAsnGluGluLeuLysValGlnPheAlaLysAsnPheValTyrAspAenAen 220  
DB 621 CCTTTGTGCAACGAGGAGATTAAGTACAAATTCGCGAAAAATTTTGTGTACGATAATAAT 680  
QY 221 SerIleLeuArgValProGluSerPheHisAspProAenArgPheGluGlnSerLeuGlu 240  
DB 681 TCAATTTCTCGAGTTCCTCGAATCGTTTACAGATCCAAACAGATTCGAAACAATCATTTAGAA 740  
QY 241 ValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyIleLysGluLeuPheAsp 260  
DB 741 GTAGCACCAAGATCGAAGCATGTTTGGAAATTTACATTTGGAATCAAAGAAATTTGTCGAT 800  
QY 261 GlyGluProValLeuAenPheAlaIleValAspLysLeuPheTyrAenAlaProLysMet 280  
DB 801 GGTGAACCTGTGCTCAATTTTGCATTTGCAATTCGATAAACTATTCTACAAATGCACCGAAAAATG 860  
QY 281 SerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSerCysAsnAspAspValArg 300  
DB 861 TCTCTTCTGGATATCTCTCTCTAAATTTGTCGACCCCGTCGTGTAAACGATGATGTACGA 920  
QY 301 LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgPro 320  
DB 921 AAAGATCTTAAACAACAACTGATGCGGGGAAAAATGACAAATCAGACAACCGCGCGGCCA 980  
QY 321 ArgIleArgGlnLeuLeuLeuLeuLysLeuLysCysAlaGluValTrpAspAenGlu 340  
DB 981 AGAATTCGACAAATTTAGAAAAATTTGAAGCTGAAATGCGCAGAAGTTTGGGATAACGAA 1040  
QY 341 MetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAenSerLeu 360  
DB 1041 ATGTGAGATTTGACAGACGACATCTGACATTTCTAGATTTTGTGCGAGGAAACTCTCTT 1100  
QY 361 ValTyrLysValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThr 380  
DB 1101 GTTTATAAGTCACCTGGTAAATCGGACAGAGGAAGAAATGCAAAAAAGTACGATACTACA 1160  
QY 381 LeuPheLysIleTyrGluGluAenLysLysPheIleGluPheProHisLeuProLeuVal 400  
DB 1161 TTGTTCAAAATCTATGAGGAAAAACAAAAAGTTCAATTGAGTTTCCCCACCTACCACTAGTC 1220  
QY 401 LysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGlu 420  
DB 1221 AAAGTTAAAGTGGAGCAAAAGAAATACGCTGTACCAATGGAAATCTTGAAGTTTCATGAG 1280  
QY 421 LysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArg 440  
DB 1281 AAGCCACAAAGATACAAAGATCGAATTTGATCTGGTGATCGAAGACAAGTTTCTAAAGCGA 1340  
QY 441 AlaThrArgLysProHisAspTyrLysGluAenThrLeuLysMetLeuLysGluLeuAsp 460  
DB 1341 GCTACACGAAAACTCACGACTACAAAGAAAAATACCCTTAAAAATGCTGAAAGAAATTTGAT 1400  
QY 461 PheSerSerGluGluLeuAenPheValGluArgPheGlyLeuCysSerLysLeuGlnMet 480  
DB 1401 TTCTCTTCTGAAGAGCTAAATTTTGTGAAGAAATTTGGAAATTTATGCTCCAAACTTCAGATG 1460  
QY 481 IleGluCysProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGln 500  
DB 1461 ATCGAATGTCAGGAAAGGTTTTGAAAGAGCCAAATGCTTGTGAATAGTGTAAATGAACAA 1520  
QY 501 IleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAenValValProGlu 520  
DB 1521 ATTTAAATGACACCAAGTATCGTGGATTTTCAAGAAAAACAAATTTGAATGTGTTCCCGAA 1580

QY 521 LysGluLeuCysCysAlaValPheValValAenGluThrAlaGlyAenProCysLeuGlu 540  
DB 1581 AAAGAACTTTGCTGTGCTGTTTTTTAGTCAACGAAACAGCGGGAATTCATGCTTAGAA 1640  
QY 541 GluAenAspValValLysPheTyrThrGluLeuIleGlyGlyCysLysPheArgGlyIle 560  
DB 1641 GAGAACGAGCTGTTTAAAGTTCTACACCGAACTAAATTTGGTGGTTCGAAGTTCCGCGAATA 1700  
QY 561 ArgIleGlyAlaAenGluAenArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAen 580  
DB 1701 CGAATTTGGTCCCAATGAAACAGAGAGCGCAATCTATTATGACACCGCGCAAAAAAT 1760  
QY 581 GluTyrAlaPheTyrLysAenCysThrLeuAenThrGlyIleGlyArgPheGluIleAla 600  
DB 1761 GAATATGCTTCTACAAAAATTTGTACACTAAATACCGAATCGGTAGATTGAAATAGCC 1820  
QY 601 AlaThrGluAlaLysAenMetPheGluArgLeuProAspLysGluGlnLysValLeuMet 620  
DB 1821 GCAACAGAAAGCGAAGATATGTTGAAACGCTCTCCCGATAAAGAACAAAAAGTCTTAATG 1880  
QY 621 PheIleIleSerLysArgGlnLeuAenAlaTyrGlyPheValLysHisTyrCysAsp 640  
DB 1881 TTCAATATCATTTCCAAACGCAACTGAATGCTTACGTTTTGTGAAACATTTATTCCGAT 1940  
QY 641 HisThrIleGlyValAlaAenGlnHisIleThrSerGluThrValThrLysAlaLeuAla 660  
DB 1941 CACACATCGGTGATGCTAATCAGCATATTACTTTCTGAACAGTCAAAAGCTTTGGCA 2000  
QY 661 SerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAen 680  
DB 2001 TCACTAAGCACGAGAAAGGATCAAAACGAATTTTCTATCAAAATGTCATTGAAAAATCAAC 2060  
QY 681 AlaLysLeuGlyGlyIleAenGlnLeuAspTrpSerGluIleAlaGluIleSerPro 700  
DB 2061 GCGAAATTAGGAGGTATTAAACAGAGCTTGACTGGTCAGAAATTCAGAAATATCACCAC 2120  
QY 701 GluGluLysGluArgArgLysThrMetProLeuThrMetTyrValGlyIleAspValThr 720  
DB 2121 GAAGAAAAAGAAAGACGAAACAAATGCCATTAATATGATATGTTGGAATGATGTAAC 2180  
QY 721 HisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValValAlaSerIleAen 740  
DB 2181 CATCAACCTCTACAGTGGAAATTTGATTTCTATAGCGCTGTAGTACCGATATCAAT 2240  
QY 741 ProGlyGlyThrIleTyrArgAenMetIleValThrGlnGluGluCysArgProGlyGlu 760  
DB 2241 CAGGTGGAACTATCTATCGAAATATGATTTGATCTCAAGAAAGATGTCTCCCGTGAG 2300  
QY 761 ArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeu 780  
DB 2301 CGTGCAGTGGCTCATGGACGGGAAAGACAGATATTTTGGAGCAAAAGTTCTGTGAATTTG 2360  
QY 781 LeuArgGluPheAlaGluAenAenAspAenArgAlaProAlaHisIleValValTyrArg 800  
DB 2361 CTCAGAGAAATTCGAGAAAAACAAACCAATTCGAGCACCGACCATATTTAGTCTATCGA 2420  
QY 801 AspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLys 820  
DB 2421 GACGAGTTAGCGATTCGAGATGCTAGCTGTAGTCATGATGAGCTTCGATCTTTTAAAA 2480  
QY 821 SerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLysTyrThr 840  
DB 2481 AGCGAAGTAAACAAATTCATGTCGAAACGGGATCGAAGATCCAGAGCGGAAGTACACG 2540  
QY 841 PheIleValIleGlnLysArgHisAenThrArgLeuLeuArgArgMetGluLysAspLys 860  
DB 2541 TTCAATTTGATTCAGAAAAAGCACAAATACACATTCGCTTCGAAGAAATGGAAGAAATAAG 2600  
QY 861 ProValValAenLysAspLeuThrProAlaGluThrAspValAlaValAlaValLys 880  
DB 2601 CAGTGGTCAATTAAGATCTTACTCTCTCTGAACAGATGTCTGCTGCTGCTGCTGTA 2660  
QY 881 GlnTrpGluGluAspMetLysSerLysGluThrGlyIleValAenProSerSerGly 900



```
Db 2661 CAATGGGAGGATATGAAGAAAGCAAGAACTGGAATGTGAACCCATATCCGGA 2720
Qy ThrThrValAspLysLeuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHis 920
Db 2721 ACAACTGTGGATAACTTATCGTTTCGAAATACAAATTCGATTTTCTTGGCATCTCAT 2780
Qy 921 HisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGly 940
Db 2781 CATGGTGTCTTGTGTATCATCTCGTCCAGGACATTACACTGTATGTATGACGATAAAGGA 2840
Qy 941 MetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCys 960
Db 2841 ATGAGCCAAAGTGAAGTCTATAAATGACCTACGACCTTGCCTTCTCTGTAGATGT 2900
Qy 961 ArgLysProLysSerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAla 980
Db 2901 CGAAACCCCATCTCGTTGCCTGTTCCGGTTCATTATGCTCATTCATGTGAAAGAGC 2960
Qy 981 LysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThr 1000
Db 2961 AAGAGCTTTATCGAACTTACAAGGAACATTACATCGGTGACTATGCACAGCCACGACT 3020
Qy 1001 ArgHisGluMetGluHisPheLeuGlnThrAsnValLysTyrProGlyMetSerPheAla 1020
Db 3021 CGACACGAAATGGAACTTTTCTCCAACTAACTGAAGTACCCCTGGAATGTGCTTCGCA 3080
```

## RESULT 3

```
US-11-144-985-2
; Sequence 2, Application US/11144985
; Publication No. US20060024798A1
; GENERAL INFORMATION:
; APPLICANT: Mello, Craig C.
; APPLICANT: Tabara, Hiroaki
; APPLICANT: Grishok, Alla
; APPLICANT: Fire, Andrew
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC
; FILE REFERENCE: INTERFERENCE
; CURRENT APPLICATION NUMBER: US/11/144,985
; CURRENT FILING DATE: 2005-06-03
; PRIOR APPLICATION NUMBER: US 09/689,992
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3227
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21)...(3080)
US-11-144-985-2
```

```
Alignment Scores:
Pred. No.: 0 Length: 3227
Score: 5349.00 Matches: 1020
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 16 Gaps: 0
```

US-10-645-746-3 (1-1020) x US-11-144-985-2 (1-3227)

```
Qy 1 MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspProGlu 20
Db 21 ATGTCTCGAAATTTCCCGAATTTGGAAAAAGGATTTTATCGTCATCTCTCGATCCGGAG 80
Qy 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysLysVal 40
```

```
Db 81 ATGAAATGGCTTGCAGGCCCACTGGTAAATTCGACGCGCAAAATCTATGAGAGAAAGTA 140
Qy 41 LeuLeuLeuValLeuLeuTyrPhePheSerSerLysValLeuTyrAspArgGluTyrTyrCdu 60
Db 141 CTTCCTTTGGTAAATTTGGTTCTCCAGCAAAATTTTACGATCGGGAATATCTACGAG 200
Qy 61 TyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLys 80
Db 201 TATGAACTGAAATGACAAAGGAAGTATTGAATAGAAAACCCAGGAAACCTTTCCCAAAA 260
Qy 81 LysThrGluLeuProLysProAspArgAlaLysLeuPheThrGlnHisLeuArgHisGlu 100
Db 261 AAGACAGAAATTCGAATTCGCGATCGTGCAGAAATCTCTTCTGGCAACATCTTCGCGCATCG 320
Qy 101 LysLysGlnThrAspPheLeuGluAspTyrValPheAspGluLysAspThrValTyr 120
Db 321 AAGAGCGAGACAGATTTTATCTCGAAGACTATGTTTTGATGAAAAGGACACTGTTTAT 380
Qy 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysValLys 140
Db 381 AGTGTGTTGTCGACTGAACACTGTGCATCAAAATGCTGGTTTCGGAGAAAGTAGTAAA 440
Qy 141 LysAspSerGluLysLysAspGluLysAspLeuLysLysLysLysLysLysLysLys 160
Db 441 AAGGATTCGGAGAAAAAAGATGAAAAGGATTTGGAGAAAAAAATCTTATACACAATGATA 500
Qy 161 LeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGluLysAspGlu 180
Db 501 CTTACCTATTCGTAATAAATTTTACCTGAATTTAGTCGAGAAAATTCGGAAAAAAGAGCAA 560
Qy 181 GluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysValArgTyrAla 200
Db 561 GAAGCGAATCGGAGTTTACAAATTCCTGAAAGATTTATGACCCAGAAAGTTCGCTACGCG 620
Qy 201 PropheValAsnGluLysValGlnPheAlaLysAsnPheValTyrAspAsnAsn 220
Db 621 CCTTTTGTGAACGAGAGATTAAGATACAAATTCGCGAAAAAATTTGTGTACGATAAAT 680
Qy 221 SerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGlnSerLeuGlu 240
Db 681 TCATTTCTGCGAGTTCTGATCGTTTCCAGATCCAAACAGATTCGAACATCAATTAGAA 740
Qy 241 ValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyLysGluLeuPheAsp 260
Db 741 GTAGCACCAAGAAATCGAAGCATGGTTTGGAAATTTACATTCGAATCAAAAGATTTGTT 800
Qy 261 GlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAlaProLysMet 280
Db 801 GGTGAACCTGTGCTCAATTTTGCATTTGTGCAATTTTCTTACCAATTCACCGAAAAATG 860
Qy 281 SerLeuLeuAspTyrLeuLeuLeuValLeuAspProGlnSerCysAsnAspAspValArg 300
Db 861 TCTCTTCTGGAATTAUCTTCTCCTAATGTGCGACCCCGCATCGTGTACGATGATACGA 920
Qy 301 LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaArgPro 320
Db 921 AAAGATCTTAAAAAAGCTGATGGCGGGAATAATGCAATTCAGACAGCGCGCGGCCA 980
Qy 321 ArgIleArgGlnLeuLeuGluAsnLeuLysCysAlaGluValTrpAspAsnGlu 340
Db 981 AGAATTCGCAATTAATTGGAAAAATTTGAAAGCTGAAATGCGCAGAGTTTGGGATAACGAA 1040
Qy 341 MetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeu 360
Db 1041 ATGTGAGATTTGACAGACGACATCTGACATTTCTAGATTTTGGCGAGGAAAACTCTCT 1100
Qy 361 ValTyrLysValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThr 380
Db 1101 GTTTATAAAGTCACCTGGTAAATCGGACAGAGGAAATGCAAAAGTACGATCTACTACA 1160
Qy 381 LeuPheLysIleTyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuVal 400
```

Db 1161 TTGTTCAAAATCTATGAGGAAACAAAGATTTTCATTGAGTTTCCCACTACCACTAGTC 1220  
Qy 401 LysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuValHisGlu 420  
Db 1221 AAAGTTAAAGTGGAGCAAAAGAAATACGCTGTACCAATCGAAATCTTGAAGTTCAAG 1280  
Qy 421 LysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArg 440  
Db 1281 AAGCCCAAGATACAAAGATTCGAATTCGATCTGGTATCGAAGCAAGTTCTTAAGCGA 1340  
Qy 441 AlaThrArgLysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAsp 460  
Db 1341 GCTACAGCAAAACCTCACGACTACAAAGAAATATCCCTAAAAATGCTGAAGAAATGGAT 1400  
Qy 461 PheSerSerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMet 480  
Db 1401 TTCTCTTCTGAAGAGCTAAATTTTGTGAAGATTTGGAATTTATGCTCCAAATCTCAGATG 1460  
Qy 481 IleGluCysProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGln 500  
Db 1461 ATCGAATGTCAGGAAGAGGTTTGAAGAGCCATGCTTGTGATGTGAATGAACAA 1520  
Qy 501 IleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGlu 520  
Db 1521 ATTTAAATGACACCACTGATTCGTGATTTCAAGAAAAACAATTTGAATGCTGCCGAA 1580  
Qy 521 LysGluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGlu 540  
Db 1581 AAAGAACTTTGCTGTCTGTTTTGTAGTCAACGAAACAGCGGAAATCCATGCTTAGAA 1640  
Qy 541 GluAsnAspValValLysPheTyrThrGluLeuIleGlyCysLysPheArgGlyIle 560  
Db 1641 GAGACGAGCTGTTGAAGTTCTACACGAACTAATTTGGTGGTGCAGTTCCGTGGAATA 1700  
Qy 561 ArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAsn 580  
Db 1701 CGAATTTGGTGCCAAATGAAACAGAGAGCGCAATCTATTATGTACGACGCGACGAAAT 1760  
Qy 581 GluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAla 600  
Db 1761 GAATATGCTCTTACAAAAAATTTGTACATAAATACCGGAATCGGTAGATTGAAATAGCC 1820  
Qy 601 AlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLysValLeuMet 620  
Db 1821 GCACAGAGCGAGAAATATGTTGACGCTCTCCGATAAGAACAAAAAGCTTTAATG 1880  
Qy 621 PheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAsp 640  
Db 1881 TTCAATATCATTTCCAAACGACAACTGAATGCTTACGGTTTTGTGAAAAATATTATTGCGAT 1940  
Qy 641 HisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAla 660  
Db 1941 CACACCATCGGTAGCTAATCAGCATATTTACTTCTGAACACAGTCAAAAGCTTTGGCA 2000  
Qy 661 SerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsn 680  
Db 2001 TCATTAAGCAGCAGAAAGATCAAAACGAAATTTCTATCAATTTGCAATGGAATCAAC 2060  
Qy 681 AlaLysLeuGlyIleAsnGlnLeuAspTyrPheSerGluIleAlaGluLysSerPro 700  
Db 2061 GCGAATATTAGGAGTATTAAACAGAGAGCTTGACTGGTCAGAAATTCAGAAATATCACCA 2120  
Qy 701 GluGluLysGluArgLysThrMetProLeuThrMetTyrValGlyIleAspValThr 720  
Db 2121 GAAGAAAAAGAAAGCGAAACAAATGCCATTAACTATGATGTGGAATTCATGTAATCT 2180  
Qy 721 HisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValValAlaSerIleAsn 740  
Db 2181 CATCCAACTCTACAGTGAATTTGATTAATCTATAGCGGCTGTAGTAGCGATCAAT 2240  
Qy 741 ProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluCysArgProGlyGlu 760  
Db 2241 CCAGGTGGAATCTATTCGAAATATGATTTGACTCAAGAGAAATGTCGTCGCCGTGAG 2300

Qy 761 ArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeu 780  
Db 2301 CGTGAGTGGCTCATGGACGGGAAAGAACAGATATTTTGGAAAGCAAGTTCTGTGAATTTG 2360  
Qy 781 LeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAlaHisIleValValTyrArg 800  
Db 2361 CTCAGAGAATTCGCAGAAAAACAACGACAAATCGAGCAGCCAGCATATTCTAGTCTATCGA 2420  
Qy 801 AspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLys 820  
Db 2421 GACGAGATTAGCGATTCGGAGATGCTGTTAGTTCATGATGAGCTTCGATCTTTAAAAA 2480  
Qy 821 SerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLysTyrThr 840  
Db 2481 AGCGAAGTAAACAATTTCTATCGGAACGGGATCGAAGATCCAGAGCCGGAAGTACAG 2540  
Qy 841 PheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgMetGluLysAspLys 860  
Db 2541 TTCAATTGTGATTCAGAAAAAGACACAATACACGATTTGCTCGAAGAAATGCAAAAGATAAG 2600  
Qy 861 ProValValAsnLysAspLeuThrProAlaGluThrAspValAlaValAlaValLys 880  
Db 2601 CCAGTGGTCAATAAAGATCTTACTCTGCTGAAACAGATGTCGCTGTTCTGCTGTATAA 2660  
Qy 881 GlnTyrGluGluAspMetLysGluSerLysGluThrGlyIleValAsnProSerSerGly 900  
Db 2661 CAATGGGAGGAGATATGAAAGAAAGCAAGAAATCGAATTTGTGAACCATCATCCGGA 2720  
Qy 901 ThrThrValAspLysLeuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHis 920  
Db 2721 ACAACTGTGGATAAACTTATCGTTTCGAAATACAAATTCGATTTTCTTGGCATCTCAT 2780  
Qy 921 HisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspLysGly 940  
Db 2781 CATGTTGCTTGGTACATCTCGTCCAGACATTACACTGTTATGTATGACGATAAAGGA 2840  
Qy 941 MetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCys 960  
Db 2841 ATGACCAAGATGAAGTCTATAAATGACCTACGACATGCTTTCTCTCTCTAGATGT 2900  
Qy 961 ArgLysProIleSerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAla 980  
Db 2901 CGAAACCCATCTCGTTCGCTTCCGTTTCATTTATGCTCATTTATCATGTGAAAAAGCG 2960  
Qy 981 LysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThr 1000  
Db 2961 AAAGAGCTTTATCGAATCTTACAGGAACATTACATCGGTGACTATGCACAGCCCGACT 3020  
Qy 1001 ArgHisGluMetGluHisPheLeuGlnThrAsnValLysTyrProGlyMetSerPheAla 1020  
Db 3021 CGACAGCAAAATGGAACATTTCTCCAACTAACGTGAAGTACCTCGGAATGCTGTTCCGA 3080

## RESULT 4

US-10-645-746-1  
; Sequence 1, Application US/10645746  
; Publication No. US20040265839A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; TITLE OF INVENTION: INTERFERENCE  
; FILE REFERENCE: US-052DVI  
; CURRENT APPLICATION NUMBER: US/10/645,746  
; PRIOR FILING DATE: 2003-08-20  
; PRIOR APPLICATION NUMBER: US 09/689,992  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/193,218  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: US 60/159,776  
; PRIOR FILING DATE: 1999-10-15

```
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3719
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-10-645-746-1

Alignment Scores:
Pred. NO.: 0 Length: 3719
Score: 5001.00 Matches: 1016
Percent Similarity: 85.2% Conservative: 1
Best Local Similarity: 85.1% Mismatches: 3
Query Match: 93.5% Indels: 177
DB: 9 Gaps: 10

US-10-645-746-3 (1-1020) x US-10-645-746-1 (1-3719)

Qy 1 MetSerSerAenPheProGluLeuGluLysGlyPheTyArgHisSerLeuAenPro--- 19
Db 21 ATGTCCTCGAATTTTCCCGAATTGGAAAAAGGATTTTATCGTCATTCTCTCGATCCGGTA 80
Qy 20 -----GluMetLy 22
Db 81 TGATCAATTATTACGACTATAGATATATAAGTTTGATTAATAATTATAGGAGTAA 140
Qy 22 sTTPLeuAlaArgProThrGlyLysCysAspGlyLysPheTyArgHisSerLeuAenPro 42
Db 141 ATGCGTTGCGAGGCCCATCTGTAATTCGACGGCAATTTCTATGAGAAGAAAGTACTTCT 200
Qy 42 uLeuValAenTTPheLysPheSerSerLysLysLysLysLysLysLysLysLysLys 62
Db 201 TTTGGTAAATTTGGTTCAAGTTCTCCAGCAAAATTTACGATCGGGAATTAATACGAGTATGA 260
Qy 62 uValLysMetThrLysGluValLeuAenArgLysProGlyLysProPheProLysLysTh 82
Db 261 AGTGAAATGACAAAGAGATTTGATATAGAAACACAGGAACCTTTCCAAAAAAGAC 320
Qy 82 rGluLeuProLe-----ProAs 88
Db 321 AGAAATTCCAATGTAAGTGTGTGAAATTTAGTCAAAACTAATTTATTTTTCAGTCCGA 380
Qy 88 pArgAlaLysLeuPheTTPGlnHisLeuArgHisGluLysLysGlnThrAspPheLeLe 108
Db 381 TCGTGCAAAACTCTTTCGCAACATCTTCGGCATTCGAGAACGACAGATTTTATTTCT 440
Qy 108 uGluAspTyrValPheAspGluLysAspThrValTyrSerValCysArgLeuAenThrVa 128
Db 441 CGAAGACTATGTTTGTGTAAGAGGACACTGTTTATAGTGTGTCGACTGAACACTGT 500
Qy 128 lThrSerLysMetLeuValSerGluLysValLysLysLysLysLysLysLysLysLys 148
Db 501 CACATCAAAAATGCTGTTTTCGGAGAAAGTAGTAAAAAAGGATTCGGAGAAAAAAGATGA 560
Qy 148 uLysAspLeuGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 168
Db 561 AAAGGATTTTCGAAAAAATCTTATACAAATGATATTCACCTATCGTAAAAAATTTCA 620
Qy 168 sLeuAenPheSerArgGluAenProGluLysAspGluLysLysLysLysLysLysLys 188
Db 621 CCTGAATTTAGTCGAGAAATTCGGAAGAAAGACGAAGAGCGAATTCGGAGTTTACAAAT 680
Qy 188 eLeuLys-----AsnValMe 193
Db 681 CCTGAA-GGTTTATGAAAAACCGCAATTATAACAAACAAATTTAGCTTTTCAGAAATGTTAT 739
Qy 193 tThrGlnLysValArgTyrAlaProPheValAenGluLysLysVal 209
Db 740 GACCCGAAAGTTTCGTACGCGCTTTTGTGAACGAGGAGATTTAAAGTGTGAGTTGCAAT 799
Qy 210 -----GlnPheAlaLysAenPh 215
Db 800 AATAATAATAATAATCACTCACTGATTTATATATTTTAAAGCAATTTCCGGAATAATTT 859

215 eValTyrAspAenAenSerLysLeuArgValProGluSerPheHisAspProAenArgph 235
860 TGTGTACGATAATAATCAATTTCTCGAGTTCTCGAATCGTTTTCAGCATCCAAACAGATT 919
235 eGluGlnSerLeuGluValAlaProArgLysGluLysLysLysLysLysLysLysLys 255
920 CGAACAAATCATTAGAGTAGCACCAAGATCGAAGCATGGTTTGGAAATTTTACATTTGAAT 979
255 eLysGluLeuPheAspGlyGluProValLeuAenPheAlaLe----- 269
980 CAAAGAATTTGTCGATGCTGACCTGCTCAATTTTTCGAAG-TAAGTTTCGAGAACTGC 1038
270 -----ValAspLysLeuPheTyArgHisSerLeuAenProLy 279
1039 GATAAAAAATCATGTGATTTTGTGAAGTTGTCGATAAACTATTTACAAATGCACCGAA 1098
279 sMetSerLeuLeuAspTyrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 299
1099 AATGTCTCTTCTCGATTTATCTTCTCTTAATTTGACCCCGCTGTCGTAAACGATGATGT 1158
299 lArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrLysArgGlnAlaAa 319
1159 ACGAAAGATCTTAAACAAACTGATGGCGGGAATAATGACAAATCAGACGAGCCGCGC 1218
319 gProArgLysArgGlnLeuLeuLeuLeuLysLysLysLysLysLysLysLysLysLys 339
1219 GCCAAGAAATTCGACAAATTTTGGAAATTTGAAAGTGAATGCGCAGAGTTTGGGATAA 1278
339 nGluMet-----SerArgL 344
1279 CGAAATGTTAGTTTAAATTTCAAACTAATAATATATAAAATTTGATTTTCAGGTCGAGAT 1338
344 euThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAenSerLeuValTyrLysV 364
1339 TGACAGACGACATCTGACATTTCTAGATTTGCGAGGAAACTCTCTCTGTTTATAAAG 1398
364 alThrGlyLysSerAspArgGlyArgAenAlaLysLysLysLysLysLysLysLysLys 384
1399 TCACGTGTAATTCGACAGAGGAGAAATGCAAAAAAGTAGTACGATCTACATTTGTTCAAAA 1458
384 lTyrGluGluAenLysLysPheLeuGluPheProHisLeuProLeuValLysLysLys 404
1459 TCTATGAGGAAAAACAAAGTTTCAATGAGTTTCCCCACCTACCACTAGTCAAAAGTTAAAA 1518
404 erGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnA 424
1519 GTGGAGCAAAAGAAATACGCTGTACCAATGGAACATCTTGTGAAGTTTCATGAGAGCCACAA 1578
424 rGtyrLysAenArgLysLeuValMetGlnAspLysPheLeuLysArgAlaThrArgL 444
1579 GATACAAAGAAATCGAAATGATCTGCTGATGCAAGCAAGTTTCTAAAGCGAGCTACACGAA 1638
444 ySProHisAspTyrLysGluAenThrLysLysMetLeuLysGluLeuAspPheSerSerG 464
1639 AACCTCAGCACTACAAAGAAATACCCCTAAAAATGCTGAAAGAAATTTGGATTTCTCTCTG 1698
464 luGluLeuAenPheValGluArgPheGlyLeuCysSerLysLysLeuGlnMetLysGluCysP 484
1699 AAGAGCTAAATTTTGTGAAAGATTTGGATTTGCTCCAACTTCAGATGATCGAATGTC 1758
484 rGlyLysValLysGluProMetLeuValAenSerValAenGluGlnLysMetT 504
1759 CAGGAAAGGTTTGAAGAGCCCAATGCTGTGTAATAGTGTAAATGAACAAATTAATAATGA 1818
504 hrProValLysArgGlyPheGlnLysGlnLeuAenValValProGluLysGluLeu 524
1819 CACCACTGATTCGTGAGTTTCAAGAAAAACAAATTTGAATGTGTTTCCCGAAAAAGAACTTT 1878
524 ySValAlaValPheValValAenGluThrAlaGlyAenProCysLeuGluGluAenAspV 544
1879 GCTGTGCTGTTTGTGATGTCACCAACACGCGGGAATAATCCATGCTTAGAAGAGACGACG 1938
```

544 alVal-----LysPheTyrThrGlu 550  
 |||||  
 1939 TTGTGTAAGTGTCTACGTAGATTATTCGGAATAATTTTCAGTAAAGTTCTACACCGAA 1998  
 551 LeuileGlyCysLysPheArgGlyileAtrGileGlyAlaAsnGluAenArgGlyAla 570  
 1999 CTAATTGGTGGTTCGAAGTTCGGTGAATACGAATGGTGGCAATGAAACACAGAGAGCG 2058  
 571 GlnSerileMetTyrAspAlaThrLysAsnGlu----- 581  
 2059 CAATCTATTATGTACACGCGACGAAATAATGAAATATGTCGCGTAAATTCAGAAATAATGAAA 2118  
 582 -----TyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyileGly 595  
 2119 GTTTTAAATATCATATTACAG-TTCTACAAAATTTGACATAAATCCGGAATCGGT 2177  
 596 ArgPheGluileAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGlu 615  
 2178 AGATTTGAAATAGCCGCAACAGAGCGAAGAAATATGTTTGAACGCTCTCCCGATAAAGAA 2237  
 616 GlnLysValLeuMetPheilelleSerLysArgGlnLeuAsnAlaTyrGlyPheVal 635  
 2238 CAAAAAGTCTTAATGTTCATATTCATTTCCAAACGCAACTGAATGCTTACCGTTTTGTG 2297  
 636 LysHisTyrCysAspHisThrileGlyValAlaAsnGlnHisileThrSerGluThrVal 655  
 2298 AAACATTTTGGATCACACCATCGGTGAGCTAATCACGATATTACTTTCGAAACAGTC 2357  
 656 ThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgilePheTyrGlnile 675  
 2358 ACAAAGCTTTGGCATCACTAAGGCACGAGAAAGGATCAAAACGAAATTTTCTATCAATT 2417  
 676 AlaLeuLysileAsnAlaLysLeuGlyLysileAsnGlnGluLeuAspTrpSerGluile 695  
 2418 GCATTTGAAATCAACCGAAATTTAGGAGGTATTAACCGAGGCTTCACTGGTCAGAAATT 2477  
 696 AlaGluileSerProGluLysGluArgLysThrMetProLeuThrMetTyrVal 715  
 2478 GCAGAAATATCACGAGAAAGAAAGAAACGGAAGAAACAAATGCCATTAACTATGATGTT 2537  
 716 GlyileAspValThrHisProThrSerTyrSerGlyileAspTyrSerileAlaAlaVal 735  
 2538 CGAATTTGATTAACCTCATCCACCTCTACAGTGGAAATTTGATTTCTATAGCGGCTGTA 2597  
 736 ValAlaSerileAsnProGlyThrileTyrArgAsnMetileValThrGlnGluGlu 755  
 2598 GTAGCGAGTATCAATCCAGGTGGAATCTATCTCGAAATATGATTTGTGACTCAAGAAGAA 2657  
 756 CysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspileLeuGluAla 775  
 2658 TGTCTGTCGGTGGAGGTGAGTGGCTCATGGACGGGAAGAACAGATATTTTGGNAGCA 2717  
 776 LysPheValLysLeuLeuArgGluPheAlaGlu----- 786  
 2718 AAGTTCGTGAAATGCTCAGAGAAATTCGCAAGATTCGCAAGATGAGTTGCTTGGATTTTAAAGAT 2777  
 787 -----AsnAsnAspAsnArgAlaProAlaHi 795  
 2778 CTCTCGGATTTTTTAATTTTTTTGTAAACTTTTTCAGAAACAACTCGAGACCGACGCGCA 2837  
 795 eilleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspG1 815  
 2838 TATTGTAGTCTATCGAGCGGAGTTAGCGATTCGGAGATGCTACGTTGATTAGTCATGATGA 2897  
 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835  
 2898 GCTTCGATCTTTAAAAAGCGAATGAAACAATTCATGTCGGAACCGGATCGGAAGATCC 2957  
 835 oGluProLysTyrThrPheileValileGlnLysArgHisAsnThrArgLeuLeuArgAr 855  
 2958 AGAGCGCGAAGTACACGTTTCAATGTGATTGAGAAAGACAAATACACGATTCCTCGAAG 3017  
 855 gMetGluLysAspLysProValValAlaAsnLysAspLeuThrProAlaGluThrAspValAl 875

3018 AATGAAAAAGATAAGCCAGTGGTCAATAAAGATCTTACTCTGCTGAAACAGATGTCGC 3077  
 875 aValAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyileVa 895  
 3078 TGTTCTGCTGTGTTAAACAATGGAGGAGGATATGAAGAAGAAACAAAGAACTGGAATTGT 3137  
 895 IasnProSerSerGlyThrThrValAspLysLeuileValSerLysValTyrLysPheAspPh 915  
 3138 GAACCATCATCCGGAACAACCTGTGGATAAATCTATCGTTTGAATAACAAATTCGATTT 3197  
 915 ePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935  
 3198 TTTCTTGGCATCTCATCTGTTGCTTGGTACATCTGTCAGGACATTTACACTGTTAT 3257  
 935 tTyrAspAspLysGlyMetSerGluAspGluValTyr----- 947  
 3258 GTATGACGATAAAGGAATGAGCAAGATGAAAGTCTATGTAAAGCGTTTGAATAGCAGTTA 3317  
 948 -----LysMetT 950  
 3318 GCGATTTAGGATTTTGTAAATCCGCATATAGTTATTATAAAAAAATGTTTCAGAAATAGA 3377  
 950 hrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProileSerLeuProValProv 970  
 3378 CCTACGGACTTGTCTTCTCTGCTAGATGTGGAACCCCATCTCGTTGCTGTTCCGG 3437  
 970 alHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluH 990  
 3438 TTCAATATGCTCATTTATCATGTGAAAAAGCGAAAGAGCTTTATCGAACTTACAAGGAAC 3497  
 990 isTyrileGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnT 1010  
 3498 ATTACATCGGTGACTATGCACAGCCAGCATCTCGACACGAAATGGAAACATTTTCTCCAAA 3557  
 1010 hrAsnValLysTyrProGlyMetSerPheAla 1020  
 3558 CTAACGTGAAGTACCCCTGGAATGTCGTGCGCA 3589

RESULT 5

US-10-645-735-1  
 ; Sequence 1, Application US/10645735  
 ; Publication No. US20050100913A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mello, Craig C.  
 ; APPLICANT: Tabara, Hiroaki  
 ; APPLICANT: Grishok, Alla  
 ; APPLICANT: Fire, Andrew  
 ; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
 ; FILE OF INVENTION: INTERFERENCE  
 ; FILE REFERENCE: UM-G-052  
 ; CURRENT APPLICATION NUMBER: US/10/645,735  
 ; CURRENT FILING DATE: 2003-08-20  
 ; PRIOR APPLICATION NUMBER: US/09/689,992A  
 ; PRIOR FILING DATE: 2000-10-13  
 ; PRIOR APPLICATION NUMBER: US 60/193,218  
 ; PRIOR FILING DATE: 2000-03-30  
 ; PRIOR APPLICATION NUMBER: US 60/159,776  
 ; PRIOR FILING DATE: 1999-10-15  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 3719  
 ; TYPE: DNA  
 ; ORGANISM: Caenorhabditis elegans  
 US-10-645-735-1

Alignment Scores:  
 Pred. No.: 0 Length: 3719  
 Score: 5001.00 Matches: 1016  
 Percent Similarity: 85.2% Conservative: 1  
 Best Local Similarity: 85.1% Mismatches: 3  
 Query Match: 93.5% Indels: 177

DB: 10 Gaps: 10  
US-10-645-746-3 (1-1020) x US-10-645-735-1 (1-3719)  
QY 1 MetSerSerAenPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspPro--- 19  
DB 21 ATGTCCTCGAATTTTCCGAATTTGGAANAAGGATTTTATCGTCATTCTCTCGATCCGGTA 80  
QY 20 -----GluMetLys 22  
DB 81 TGATCAATTATTAGCAGCTATAAGATATATAAGTTTTCATATTAATATTATAGGAGATGA 140  
QY 22 sTcPLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysLysValLeuLe 42  
DB 141 ATGGCTTCGAGGCCCTCTGGTAATTCGACGCCAATTTCTATGAGGAAGAAGTACTTCT 200  
QY 42 uLeuValAsnTrpPheLysPheSerLysIleTyrAspArgGluTyrGluTyrG1 62  
DB 201 TTTGGTAAATTTGGTTCAAGTTCTCCAGCAAAATTTACGATCGGAATACTACGAGTATGA 260  
QY 62 uValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysTh 82  
DB 261 AGTGAATAATGACAAAGGAATTTGATAGAAAACCAAGAAAACCTTTCCCAAAAAGAC 320  
QY 82 rGluIleProIle-----ProAs 88  
DB 321 AGAATTTCCAAATGTAAGTCTGTAAATTAAGTCAAAACTAATTTATTTTTCAGTCCCGA 380  
QY 88 pArgAlaLeuPheTrpGlnHisLeuArgHisGluLysGlyGlnThrAspPheIleLe 108  
DB 381 TCGTGCAAAACCTCTCTGCGCAACATCTTCGCGCATGAGAGAAGACGACAGATTTATCT 440  
QY 108 uGluAspTyrValPheAspGluLysAspThrValTyrSerValCysArgLeuAsnThrVa 128  
DB 441 CGAAGACTATGTTTTTGTGATGAAAAGGACACTGTTTATAGTGTGTTGTCGACTGAACACTGT 500  
QY 128 lThrSerLysMetLeuValSerGluLysValValLysLysAspSerGluLysLysAspG1 148  
DB 501 CACATCAAAANATCTGTTTCGGAGAAAGTAGTAAANAAGANTTCGGAGAAAAGAGATGA 560  
QY 148 uLysAspLeuGluLysLysIleLeuTyrThrMetIleLeuThrTyrArgLysLysPheHi 168  
DB 561 AAAGGATTTGGAGAAAAAATCTTATACACATGATACTTACCTATCGTAAANAATTTCA 620  
QY 168 sLeuAsnPheSerArgGluAsnProGluLysAspGluGluLysValLysAsnArgSerTyrLysPh 188  
DB 621 CCTGAACCTTTAGTCGAGAAAAATCCGGAANAAGACGGAAGCGAATCGGAGTTTACAAAT 680  
QY 188 eLeuLys-----AsnValMe 193  
DB 681 CCTGAA-GGTTTATGAAAACACGCAATTATTAACAAATAATTTAGCTTTTCAGAAATGTTAT 739  
QY 193 tThrGlnLysValArgTyrAlaProPheValAsnGluGluIleLysVal----- 209  
DB 740 GACCCAGAAAGTTCTGCTACGCGCTTTTGTGAACGAGGAGATTAAGTGTGAGTTGCAAT 799  
QY 210 -----GlnPheAlaLysAsnPh 215  
DB 800 AATAATAATAATAATCACCTCAACTCATTTATATATTTTAAAGACAATTCGCGAAAAATTT 859  
QY 215 eValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPh 235  
DB 860 TGTGTACGATAATAATTTCAATCTCGGAGTTCCTGAATCGTTTCACGATCCAAACAGATTT 919  
QY 235 eGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyI1 255  
DB 920 CGAACATCATTAGAAGTAGCACCAGAATCGAAGCATGGTTTGGAAATTTACATTCGAAT 979  
QY 255 eLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIle----- 269  
DB 980 CAAAGAATTTGTCGATGGTGAACCTGCTGCTCAATTTTGCAAG-TAAGTTTGAGAAACTGC 1038  
QY 270 -----ValAspLysLeuPheTyrAsnAlaProLys 279

DB 1039 GATAAAAAATCATGTGTGATTTTGTGAAGTTGCGATANAATAATTCTTACAAATGACCGAA 1098  
QY 279 sMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSerCysAsnAspAspVa 299  
DB 1099 AATGTCTCTCTCGATTAATCTTCTCTAAATTTGCGACCCCGCATGCTGTGTAAACATGATGT 1158  
QY 299 lArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaLaAr 319  
DB 1159 ACGAAAAGATCTTTAAACCAAAACTGATGCGCGGAAAAAATGACAAATCAGACAACGCGCG 1218  
QY 319 gProArgIleArgGlnLeuLeuGluAsnLeuLysCysAlaGluValTrpAspAs 339  
DB 1219 GCCAAGAAATTCGACAATTTATGGAAAATTTGAAGCTGAAAATGCGCAGAAGTTTGGGATAA 1278  
QY 339 nGluMet-----SerArgL 344  
DB 1279 CGAAATGTTAGTTTAAATTTATTAATATACAAATTAATGATTTTTCAGGTCGAGAT 1338  
QY 344 euThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysV 364  
DB 1339 TGACAGAACGACATCTGACATTTCTAGATTTGTGCGAGGAAAACCTCTCTGTTTATATAAG 1398  
QY 364 alThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysI 384  
DB 1399 TCACCTGGTAAATCGGACAGAGGAAGAAATGCAAAAAGTACGATACTACATTTGTTCAAAA 1458  
QY 384 lTyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLys 404  
DB 1459 TCTATGAGGAACAACAAAAGTTTCATTTAGTTCCTCCACCTACCACTAGTCAAGTTAAAA 1518  
QY 404 erGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnA 424  
DB 1519 GTGAGCAAAAGAAATACGCTGTACCAATGGAACATCTTGAAGTTTCATGAGAAGCACAAA 1578  
QY 424 xGTrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysAcrAlaThrArgL 444  
DB 1579 GATCAAGAAATCGAATTTGATCTGGTGTGCAACAAAGTTTCTTAAAGCGAGCTACACGAA 1638  
QY 444 ySProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLysAspPheSerSerG 464  
DB 1639 AACCTCACGACTACAAAGAAATACCTTAAANAATGCTGAAAGAAATTCGATTTCTCTCTG 1698  
QY 464 luGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLysLeuGlnMetIleGluCysP 484  
DB 1699 AAGAGCTAAATTTTGTGAAAGATTTGGATTTATGCTCCAAACTTCAGATCATCGAATGTC 1758  
QY 484 xGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetT 504  
DB 1759 CAGGAAAGGTTTGAAGAGAGCCCAATGCTGTGAATAGTGTAAATGAACAAATTAATAATGA 1818  
QY 504 hrProValIleArgGlyPheGlnLysGlnLeuAsnValValProGluLysGluLeuC 524  
DB 1819 CACCACTGATTCGTGGATTTCAAGAAAACAAATTAATGATGTTGCTCCGAAAAAGAACTTT 1878  
QY 524 ySAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAspV 544  
DB 1879 GCTGTGCTGTTTGTGTAGTCAACGAAACAGCGGAAAAATCCATGCTTAGAAGAAACGACG 1938  
QY 544 alVal-----LysPheTyrThrGlu 550  
DB 1939 TGTGTAAAGTGTTCCTACGTAGATTAATTCGGAATAATTTTTCAGTAAGTTCTACACCGAA 1998  
QY 551 LeuIleGlyGlyCysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyVala 570  
DB 1999 CTAAATTTGGTGGTTGCAAGTTCCGTGGAATACGAATTTGGTCCCAATGAAACACAGAGCG 2058  
QY 571 GlnSerIleMetTyrAspAlaThrLysAsnGlu----- 581  
DB 2059 CAATCTATTATGTACGACGCGCAAAAAATGAATATGCGCGTAAGTTTTCAGAAAAATGAAA 2118  
QY 582 -----TyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGly 595

Db 2119 GTTTTTAAATATCATATTATACAG-TTCTACAAAAATTGTACACTAAATACCGAATCGGT 2177  
QY 596 ArgPheGluIleAlaIaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGlu 615  
Db 2178 AGATTTTGAATAGCCGCAACAGAGCGAAGAAATATGTTTGAACGCTCTCCCGATGAAGAA 2237  
QY 616 GlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaLysGlyPheVal 635  
Db 2238 CAAAAAGTCTTAATGTTTCATATATTCATTTCCAAACGACACTGAATGCTTACCGTTTGTG 2297  
QY 636 LysHisTyrCysAspHisThrIleGlyValAlaLysGlnHisIleThrSerGluThrVal 655  
Db 2298 AAACATTTATGCGATCACACCATCGGTGTAGCTAATCAGCATATATCTTCTGAAACAGTC 2357  
QY 656 ThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIle 675  
Db 2358 ACAAAGAGCTTTGGCATCACTAAGCGACGAGAAAGGATCAAAACGAAATTTTCTATCAAAAT 2417  
QY 676 AlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTrpSerGluIle 695  
Db 2418 GCATTGAAATCAACCGAAATTTAGGAGGTATTAACGAGAGCTTGACTGGTCAGAAAT 2477  
QY 696 AlaGluIleSerProGluLysGluArgArgLysThrMetProLeuThrMetTyrVal 715  
Db 2478 GCAGAAATATCACCAAGAAAGAAAGACGCGAAACCAATGCCATTAATCTATGTATGTT 2537  
QY 716 GlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaVal 735  
Db 2538 GGAATTTGATGTAACTCATCCACCTCCTACAGTGGAAATTTGATTTATCTATAGCGGCTGA 2597  
QY 736 ValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGlu 755  
Db 2598 GTAGCGAGTATCAATCCAGGTGNACTATCTATCGAAATATGATGTGACTCAAGAAGAA 2657  
QY 756 CysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAla 775  
Db 2658 TGTGTCCTCCGTCGAGCGTCAGTGGCTCATGGACGGGAAAGAACAGATATTTTGAAGCA 2717  
QY 776 LysPheValLysLeuLeuArgGluPheAlaGlu----- 786  
Db 2718 AAGTTGCGTGAATTTGCTCAGAGAAATTCGCAAGATGAGTTGCTTTGAGTATTTAAAGAT 2777  
QY 787 -----AsnAsnAspAsnArgAlaProAlaHis 795  
Db 2778 CTCCTGGATTTTAAATTTTTTGTAAACCTTCGAAACACAGCAATCGAGCACCGCGCA 2837  
QY 795 sIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspG1 815  
Db 2838 TATTGTAGTCTATCGAGACGAGTTAGCGATTCCGAGATGCTACGTGTTAGTCATGATGA 2897  
QY 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835  
Db 2898 GCTTCGATCTTTAAAGCGAAGTAAACAAATTCATGTCGGAACCGGATGGAGAGATCC 2957  
QY 835 oGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgAr 855  
Db 2958 AGAGCCCAAGTACACGTTTCAATGTGTGATTCAAGAAACACAAATACAGATTGCTTCGAG 3017  
QY 855 gMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAl 875  
Db 3018 AATGGAAAAAGATAAGCCAGTGGTCAATAAAGATCTTACTCTGCTGAAACAGATGTCG 3077  
QY 875 aValAlaIaValLysGlnTrpGluLysAspMetLysGluSerLysGluThrGlyIleVa 895  
Db 3078 TGTGTGCTGTTTAAACAATGGAGAGGATATGAAGAAAGACAAAGAACTGGAATGCT 3137  
QY 895 lAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPh 915  
Db 3138 GAACCATCATCCGAAACAACTGTGGATAAACTTATCGTTTCGAAATACAAATTCGATTT 3197  
QY 915 ePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935  
Db 3198 TTTCTTGGCATCTCATCATGGTGTCTTGGTACATCTCGTCCAGGACATTACACTGTTAT 3257

QY 935 tTyrAspAspLysGlyMetSerGlnAspGluValTyr----- 947  
Db 3258 GTATGACGATAAAGGAATGAGCCAAAGATGAAAGTCTATGTAAGCGTTTGAATGACGATTA 3317  
QY 948 -----LysMet 950  
Db 3318 GCGATTTTAGGATTTTGTAAATCCGCATATAGTTATATATAAAAAATGTTTCAGAAATGA 3377  
QY 950 hrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysPheProIleSerLeuProValPro 970  
Db 3378 CCTACCGACTTCTCTCTCTCTAGATGTGCAAAACCCCATCTCGTTCCTGTTCCGG 3437  
QY 970 alHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluH 990  
Db 3438 TTCATTATGCTCATTTATCATGTGAAAGCCGAAAGAGCTTTATCGAACTTACAAAGAAC 3497  
QY 990 lTyrIleGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnT 1010  
Db 3498 ATTACATCGGTGACTATGCACAGCCACCGACTCGACACGAAATGGAACATTTTCTCCAA 3557  
QY 1010 hrAsnValLysTyrProGlyMetSerPheAla 1020  
Db 3558 CTAACGTGAAGTACCCTCGAATGTCGTTCCGA 3589

## RESULT 6

US-11-144-985-1  
; Sequence 1, Application US/11144985  
; Publication No. US20060024798A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; FILE REFERENCE: UMY-052CN  
; CURRENT APPLICATION NUMBER: US/11/144,985  
; PRIOR FILING DATE: 2005-06-03  
; PRIOR APPLICATION NUMBER: US 09/689,992  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/193,218  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: US 60/159,776  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3719  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
US-11-144-985-1

Alignment Scores:  
Pred. No.: 0 Length: 3719  
Score: 5001.00 Matches: 1016  
Percent Similarity: 85.2% Conservative: 1  
Best Local Similarity: 85.1% Mismatches: 3  
Query Match: 93.5% Indels: 177  
DB: 16 Gaps: 10

US-10-645-746-3 (1-1020) x US-11-144-985-1 (1-3719)

QY 1 MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspPro--- 19  
Db 21 ATGTCTCGAATTTTCCGAAATTGGAAAAGGATTTTATCGTCATCTCTCGATCCGGTA 80  
QY 20 -----GluMetLy 22  
Db 81 TGATCAATTTATAGCAGCTATAGATATATATAGTTGATTAATATATAGAGATGAA 140  
QY 22 sTrPLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysValLeuLe 42



Db 141 ATGGCTTGCAGGCCACCTGGTAAATCGGACGCGCAAAATCTATCAGAGAAAGTACTTCT 200  
Qy 42 uLeuValasnTpphelysPheSerLysIleTyAspArgGluTyTyGluTyG1 62  
Db 201 TTTGGTAAATGGTTCAAGTCTCCAGCAAAATTTACGATCGGAAATPACTACGAGTATGA 260  
Qy 62 uValysMetThrLysGluValLeuAenArgLysProGlyValProPheProLysLysMh 82  
Db 261 AGTGAAATGCAAGAGAGATTTGATAGAAACCCAGGAAACCTTTCCAAAAAAGAC 320  
Qy 82 rGluileProfile-----ProAs 88  
Db 321 AGAAATCCAAATGTAAGTGTGTAAATAGTCAAAACTAATTTATTTTTCAGTCCCGA 380  
Qy 88 pArgAlaLysLeuPheTgPnHisLeuArgHisGluLysGlnThrAspPheIleLe 108  
Db 381 TCGTGCAAAATCTCTTCGGCAACATCTTCGGCATGAGAGAGACAGACAGATTTTATCT 440  
Qy 108 uGluAapTyValPheAspGluLysAspThrValTySerValCysArgLeuAenThrVa 128  
Db 441 CGAAGACTATGTTTTGATGAAAGGACACTGTTTATAGTGTGTCGACTGAACACTGT 500  
Qy 128 lThrSerLysMetLeuValSerGluLysValValLysLysAspSerGluLysLysAspG1 148  
Db 501 CACATCAAAAATCTGTTTCGGAGAAAGTAGTAAAGAGGATTCGGAGAAAGAAAGATGA 560  
Qy 148 uLysAspLeuGluLysLysIleLeuTyThrMetIleLeuThrTyArgLysLysPheHi 168  
Db 561 AAAGGATTTGGAGAAAAAATCTTATACACAATGACTTACCTATCGTAAAAAATTTCA 620  
Qy 168 sLeuAenPheSerArgGluAenProGluLysAspGluGluAlaAenArgSerTyLysph 188  
Db 621 CCTGAATCTTAGTCGAGAAATCCGGAAGAAAGACGAGAGCAATCGGAGTTTACAAAT 680  
Qy 188 eLeutys-----AenValMe 193  
Db 681 CCTGAA-GGTTTTATGAAAAACACGCATTATTAACAAACAAAATTTAGCTTTCAGAACTGTAT 739  
Qy 193 tThrGlnLysValArgTyAlaProPheValAenGluLysIleLysVal----- 209  
Db 740 GACCCAGAAAGTTTCGTACGCGCTTTTGTGAAACGAGGAGATTTAAAGTGTGAGTTGCAAT 799  
Qy 210 -----GlnPheAlaLysAsnph 215  
Db 800 AATAATAATAATACCTCAACTCATTATATATATTTTAAAGCAATTCGCGAAAAATTT 859  
Qy 215 eValTyAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAenArgph 235  
Db 860 TGTGTACGATAATAATTCATCTCGAGTTCCTGAATCGTTTCAGATCCAAACAGATT 919  
Qy 235 eGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyIleGlyI 255  
Db 920 CGAACCAATCATTAGAAGTAGCACCAAGAATCGAAGCATGTTTGGAAATTTACATTCGAAT 979  
Qy 255 eLysGluLeuPheAspGlyGluProValLeuAenPheAlaIle----- 269  
Db 980 CAAGAATTTGTCGATGGTGAACCTGTCTCAATTTTGCAAG-TAAGTTTGAGAAACTGC 1038  
Qy 270 -----ValAspLysLeuPheTyAsnAlaProLy 279  
Db 1039 GATAAAAAATCATGTGATTTTGTGAACTTGTTCGATAAATCTATCTACAAATGCCCGAA 1098  
Qy 279 sMetSerLeuLeuAspTyLeuLeuLeuIleValAspProGlnSerCysAsnAspVa 299  
Db 1099 AATGTCCTCTTCGATTATCTCTCCTAAATTTGCGACCCCACTCGTGTAAACGATGATCT 1158  
Qy 299 lArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaR 319  
Db 1159 AGAAAGATCTTTAAACAAAATCTGATGGCGGAAAAATGACAAATCAGAACAGCCGCGC 1218  
Qy 319 gProArgIleArgGlnLeuLeuGluAenLeuLysLysCysAlaGluValTyAspAs 339  
Db 1219 GCCAAGATTCGACAAATTTATGGAAATTTTGAAGCTGAATTCGGCAGAAATTTGGATTA 1278

Qy 339 nGluMet-----SerArgL 344  
Db 1279 CGAAATGTTAGTTTAAATTTATTCAAACAATTAATATACAAATGATTTTCAGGTCGAGAT 1338  
Qy 344 euThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAenSerLeuValTyLysV 364  
Db 1339 TGACAGAACACATCTGACATTTCTAGATTTGTCGAGAGAAACTCTCTCTGTTTATTAAG 1398  
Qy 364 alThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyAspThrThrLeuPheLysI 384  
Db 1399 TCACTGGTAAATCGGACAGAGAGAAATGCAAAAAGTACGATCTACTATTGTTTCAAAA 1458  
Qy 384 leTyrgluGluAenLysLysPheIleGluPheProHisLeuProLeuValLysValLys 404  
Db 1459 TCTATGAGGAAAAACAAAAGTTTCATTGAGTTTCCCACTACCACTAGTCAAAAGTTAAAA 1518  
Qy 404 eGlyAlaLysGluTyAlaValProMetGluHisLeuGluValHisGluLysProGlnA 424  
Db 1519 GTGGAGCAAAAGAAATACGCTGTACCAATGGAAATCTTTGAAGTTTCATGAGAGCCACAAA 1578  
Qy 424 rGTyLysAenArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgL 444  
Db 1579 GATACAGAAATCGAATTTGATCTGTCATGCAAGCAAGTTTCTTAAGCGAGCTACACGAA 1638  
Qy 444 ySProHisAspTyLysGluAenThrLysMetLeuLysGluLeuAenPheSerSerG 464  
Db 1639 AACCTCAGCACTACAAAGAAAAATACCTTAAAAATGCTGAAAGAAATTCGATTTCTCTCTG 1698  
Qy 464 luGluLeuAenPheValGluAargPheGlyLeuCysSerLysLeuGlnMetIleGluCysP 484  
Db 1699 AGAGCTTAAATTTTGTGTGAAAGATTTGGATTTATGCTCCAAACTTCAGATGATCGAATGTC 1758  
Qy 484 roGlyLysValLeuLysGluProMetLeuValAenSerValAenGluGlnIleLysMetT 504  
Db 1759 CAGGAAAGTTTGAAGAGCCAAATGCTTGTGAATAGTGTAAATGAACAAATTAATAATGA 1818  
Qy 504 hrProValIleArgGlyPheGlnGluLysGlnLeuAenValValProGluLysGluLeu 524  
Db 1819 CACCAGTGTGTCGATTTCAAGAAAAACAAATGAATGTGTGTCGCGAAAAAGAACTTT 1878  
Qy 524 ySCysAlaValPheValAenGluThrAlaGlyAsnProCysLeuGluGluAenAspV 544  
Db 1879 GCTGTGCTCTTTTGTAGTCAACGAAACAGCGGAAATCCATGCTTAGAAGAACGACG 1938  
Qy 544 alVal-----LysPheTyThrGlu 550  
Db 1939 TTGTGTAACTGTTTCTACGTAGATTATTCGAAATATTTTCAGTAGTTTCTACCGAA 1998  
Qy 551 LeuIleGlyGlyCysLysPheArgGlyIleArgIleGlyAlaAenGluAenArgGlyAla 570  
Db 1999 CTAATTTGGTGTTCGAAGTTCCGTGGAATACGAATTTGGTCCCAATGAACAGAGAGCG 2058  
Qy 571 GlnSerIleMetTyAspAlaThrLysAenGlu----- 581  
Db 2059 CAATCTATTATGTACGACGCGCAAAAATGAATATGCGGTAGATTTTCAGAAATTTGAAA 2118  
Qy 582 -----TyralaPheTyLysAsnCysThrLeuAenThrGlyIleGly 595  
Db 2119 GTTTTTAAATATCATATTTTACAG-TTCTACAAAATTTGTACACTAATACCGAATTCGGT 2177  
Qy 596 ArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGlu 615  
Db 2178 AGATTTGAATAGCCCAACAGAGCGAAGAAATATGTTTGAACGTCTTCCGATTAAGAA 2237  
Qy 616 GlnLysValLeuMetPheIleIleSerLysArgGlnLeuAenAlaTyGlyPheVal 635  
Db 2238 CAAAAGTCTTAAATGTTTATTATCATTTTCCAAACGCAACTGAATGCTTACGTTTGTG 2297  
Qy 636 LysHisTyCysAspHisThrIleGlyValAlaAenGlnHisIleThrSerGluThrVal 655  
Db 2298 AAACATTTTGGCATCACACCATCGGTGTAGCTTATCAGCATATTAATCTTGAACAGTC 2357



```
QY 656 ThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIle 675
Db 2358 ACAAAAGCTTTGGCATCTACTAAGGCGACGAGAAAGGATCAAAACGAAATTTCTATCAAAAT 2417
QY 676 AlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTyrSerGluIle 695
Db 2418 GCATTGAAATCAACCGGAATTAAGAGGTATTAACCGAGGCTTGACTGGTCAGAAAT 2477
QY 696 AlaGluIleSerProGluLysGluArgLysThrMetProLeuThrMetTyrVal 715
Db 2478 GCAGAAATATCACCAAGAAAGAAAGACGCGAAACAATGCCATTAACTATGTATGTT 2537
QY 716 GlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaVal 735
Db 2538 GGAATTGATGAACATCAACCTCTACAGTGGAAATGATTTATCTATACGGCTGTA 2597
QY 736 ValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGlu 755
Db 2598 GTAGCGAGTATCAATCCAGTGGACTATCTACGAAATATGATTTGTGACTCAGAGAA 2657
QY 756 CysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAla 775
Db 2658 TGTGCTCCGCTGAGCGTGCAGTGGCTCATGGACGGGAAAGAACAGATATTTTGGAGCA 2717
QY 776 LysPheValLysLeuLeuArgGluPheAlaGlu----- 786
Db 2718 AAGTTCGTGAAATGCTCAGAGAAATTCGCAGAAAGTGAGTTGCTTGAGTATTTAAAGAT 2777
QY 787 -----AenAsnAspAsnArgAlaProAlaHi 795
Db 2778 CTCTGGATTTTAATTTTTTTGTAACCTTCAGAACACAGCAATCGAGCACCGGCA 2837
QY 795 sIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspG1 815
Db 2838 TATTGTAGTCTATCGAGACGAGTTAGCGATTTCGGAGATGCTACGTTGTAGTCATGATGA 2897
QY 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835
Db 2898 GCTTCGATCTTTAAAGCGAAGTAAACAATTCATGTCGGAACGGGATGGAGAGATCC 2957
QY 835 oGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgAr 855
Db 2958 AGAGCGCAAGTAGACGTTCACTGTGATTCAGAAAGACACATACACGATGCTTCGAAG 3017
QY 855 gMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAl 875
Db 3018 AATGGAAGAAAGATAAGCCAGTGGTCAATAAAGATCTTACTCTGCTGGAACAGATGTCG 3077
QY 875 aValAlaAlaValLysGlnTrpGluLysPheMetLysGluSerLysGluThrGlyIleVa 895
Db 3078 TGTTCGCTGCTGTTAAACAATGGAGAGGATATGAAGAAAGCAAGAACTGGAAATGT 3137
QY 895 lAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPh 915
Db 3138 GAACCATCATCCGGAACAACGTGGATAAATCTATCGTTTCGAAATACAAATTCGATTT 3197
QY 915 ePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935
Db 3198 TTTCTTGGCATCTCATCATGGTGTCTTGGTACATCTCGTCAGGACATTTACACTGTTAT 3257
QY 935 tTyrAspAspLysGlyMetSerGlnAspGluValTyr----- 947
Db 3258 GTATGACGATAAAGGAATGAGCAAGATGAAGTCTATGTAAGCGTTTGAATAGCAGTTA 3317
QY 948 -----LysMetI 950
Db 3318 GCGATTTTAGGATTTTGTAAATCGCATATAGTTATTTATAAAAAATGTTTTCAGAAATGA 3377
QY 950 hTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValProv 970
Db 3378 CCTACGGAATCTGTTTCTCTCTGCTAGATGTCGAAACCCCATCTCGTTGCTGTTCCGG 3437
QY 970 aHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluH 990
```

```
Db 3438 TTCATTATGCTATTATTCATGTGAAAAGCGAAAGAGCTTTATCGAATTCACGAAC 3497
QY 990 lserTyrIleGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnT 1010
Db 3498 ATTACATCGGTGACTATGCACAGCCAGGACTCGACACGAAATCGAACATTTTCTCCAA 3557
QY 1010 hrAsnValLysTyrProGlyMetSerPheAla 1020
Db 3558 CTAACGTGAAGTACCCTCGAATGTCGTTGCA 3589

RESULT 7
US-10-007-078-3
; Sequence 3, Application US/10007078
; Publication No. US20030105042A1
; GENERAL INFORMATION:
; APPLICANT: Donna T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF EIF2C1 EXPRESSION
; FILE REFERENCE: RTS-0236
; CURRENT APPLICATION NUMBER: US/10/007,078
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 7478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (214)...(2787)
US-10-007-078-3

Alignment Scores:
Pred. No.: 1.13e-41 Length: 7478
Score: 577.00 Matches: 237
Percent Similarity: 40.2% Conservative: 173
Best Local Similarity: 23.2% Mismatches: 370
Query Match: 10.8% Indels: 243
Db: 6 Gaps: 38

US-10-645-746-3 (1-1020) x US-10-007-078-3 (1-7478)
QY 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGlu----- 37
Db 209 ATGGATCGAA- GCGGACCTTCGGGACGACGTCGGGGCTTACTGCCCCCTGCGAG 267
QY 38 -----LysLysValLeuLeu 42
Db 268 CAGGTGTTCCAGGCACCTCGCGGCTGGCATTGGCACTGTGGGAAACCAATCAAGCTC 327
QY 43 LeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGluTyrGlu 62
Db 328 CTGGCCAATTACTTTGAGGTGGACATCCCTAAGATCGACGTGTACCTACGAGGTGGAC 387
QY 63 ValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysThr 82
Db 388 ATCAGCCGGATAG- -----TGTCCCGTAGAGTC 417
QY 83 GlutIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGluLysLys 102
Db 418 AACCGGAAGTGGTGAA-----TACATGCTCAGCATTTCAAGCCT 459
QY 103 GlnThrAspPheIleLeuGluAspTyr-----ValPheAspGluLysAspThrValTyr 120
Db 460 CAG-----ATCTTTGGTGATCGCAAGCCTGTGTATGATGGAAAGAAAGAACATTTTAC 510
QY 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysVal----- 138
Db 511 -----ACTGTACAGCATGCCCCATGGCAACGAGCGGTGCGATTT 552
QY 139 ---ValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeuTyr 157
Db 553 GAGGTGACAAATCCCTCGGGGAAGGAGAT---CGAATCTTTAAGGTCTCCATCAAGTGG 609
```

158 ThrMetIleLeuThrTrpArgLysPheHisLeuAsnPheSerArgGluAsnProGlu 177  
160 CTAGCCATTGTGAGCTGGCGAATCTGTCATGAGCCCTCTGTGTCAGCGCCAGATCCCTGTT 669  
178 LysAspGluGluAlaAsnArgSerTrpLysPheLeuLysAsnValMetThrGlnLysVal 197  
179 CCCTTGAGTCTGTGCAAGCCCTGGATGTGGCCATGAGGCACCTGGCATCC-----ATG 723  
198 ArgTrpAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTrp 217  
200 AGGTACACCCCTGTG----- 738  
218 AsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln 237  
219 ---GGCGCTCTCTTCTCAGCGCTGAGGCTACTACACCCG----- 780  
238 SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTrpIleGlyIleLysGlu 257  
240 ---CTGGGGGTGGCGCAAGTCTGGTTCGGCTTTCACAGTCTGTGCGCCT 831  
258 LeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTrpAsnAla 277  
260 ---TGGAAGATGATGCTCAACATTGATGCTCAGCCACTGCTTTTATAAGCA 888  
278 ProLysMetSerLeuLeuAspTrpLeuLeuLeuLeuValAspProGlnSerCysAsnAsp 297  
280 ---CAGCCAGTGATTGATGTTTCATGTGTGAGTGTGGACATCAGGAACATAGATGAG 942  
298 AspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAla 317  
300 ---CAGCCCAAG-----CCCTCAGGACTCT 966  
318 AlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLys-----CysAla 334  
320 ---CAGCGCTTCGCTTCAACAGGAGATCAAGGCTGAGGTGGAAGTCAACCCACTGTGA 1026  
335 GluValTrpAsnAsn-----GluMetSerArgLeuThrGluArgHisLeuThr 350  
351 PheLeuAspLeuCysGluGluAsnSerLeuValTrpLysValThrGlyLysSerAspArg 370  
371 GlyArgAsnAlaLysLysTrpAspThrThrLeuPheLysIleTrpGluGluAsnLysLys 390  
391 Phe---IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTrp 409  
410 AlaValProMetGluHisLeuGluValHisGluLysProGlnArgTrpLysAsnArgIle 429  
429 TACCTTCCCTAGAGTCTGTACAT---GTGGCTGGGCGAGTGTATTAAGAAGCTG 1275  
430 AspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTrpLys 449  
449 ACCGACACACAGACCTCGACCATGATAAGCCACAGCTAGATCCGCTCCAGACAGACAG 1335  
450 GluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheVal 469  
470 GAGGAGATCAGTCCCTGTGTAAGAATGCCAGTACCACTTAGATCC-----TACATC 1389  
470 GluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys 489  
490 CAGGAATTTGGGATCAAGTGAAGATGATCATCGGAGGTGACAGGGCGAGTGTGCGCG 1449  
490 GluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro----- 505  
490 ---CAGCCCATCTTGCAAGTACGGCGCGCGGCAATTCGCCACACCCCATCAGGTGTC 1509

506 ---ValIleArgGlyPheGlnGlnLysGlnLeu---AsnValValProGluLysGluLeu 523  
510 TGGGACATGGGGGG-----AAACAGTTCTTACAAATGGATTTGAGATCAAAAGTCTGG 1560  
524 CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp 543  
544 GCATCGCTTGTGCGACCCCAAAA-----CAGTGTGCAAGAG----- 1602  
544 ValValLysPheTrpThrGluLeuIle-Gly-----GlyCysLysPhe 557  
557 GTGCTCAAGAACTTCACAGACAGCTGCGGAAGATTTTCCAAGGATGCGGGATGCCATC 1662  
557 eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTrpAspAl 577  
577 CAGGGTCAACCTTTGTTTCTGCAAAATATGCACAGGGGCA-GACACGCTG----- 1710  
577 aThrLysAsnGluTrpAlaPheTrpLysAsnCysThrLeuAsnThrGlyIleGlyArgPhe 597  
597 --- 1710  
597 eGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLys 617  
617 ---GAGCCTATGTTCCGGCATCTCAAGAACACCTACTCAGG 1748  
617 sValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTrpGlyPheValLysHis 637  
637 GCTGCAGCTCATTTGTCATCTCTGCCAGGAAGACGCCGCTGTATCTCTGAGGTGAACG 1808  
637 sTrpCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLys 657  
657 TGTCGAGATACACTCTTGGGAATGCTACGAGTGTGTGAGTGAAGAACGTGGTCAA 1868  
657 sAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTrpGlnIleAlaLe 677  
677 ---GACC-----TCACCTCAGACTCTGTCACCACTCTGCCT 1901  
677 LysIleAsnAlaLysLeuGlyIleAsnGlnGluLeuAspTrpSerGluIleAlaGln 697  
697 CAAGATCAATGTCAAACTTGTGTGCATTAACAACATCCTA----- 1941  
697 LysSerProGluGluLysGluArgLysThrMetProLeuThrMetTrpValGlyIle 717  
717 ---GTCCACACAGCGCTCTCGCTTTTCAACAGCCAGCTG---ATATTCTGGGAGC 1994  
717 eAspValThrHisProThrSerTrpSerGlyIleAspTrpSerIleAlaValAlaVal 737  
737 AGATGTTTACACACCCCGCCAGCGGATGGGAAAAAACCTTCTATCACAGCAGTGTAGG 2054  
737 aSerIleAsnProGlyGlyThrIleTrpArgAsnMetIleValThrGlnGluGluCysArg 757  
757 CAGTATGATGTCACCCAGCGGATGCTGTCTGTCTGTGCGGGTACAG----- 2103  
757 gProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGlu----- 775  
775 ---CGACCGCGCAAGAGATCAATGAAGACTTGTGC 2135  
775 aLysPheValLys---LeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAl 794  
794 ---CTACATGGTGGTGAGTCTCTCATCCAAATCTCAAGTCCACCCGTTTCAAG---CCTAC 2192  
794 aHisIleValValTrpArgAspGlyValSerAspSerGluMetLeuArgValSerHisAs 814  
814 CCGCATCATCTTCTACCGAGATGGGTGCTGCAAGGCCAGCTACCCCGAGATCTCCATTA 2252  
814 pGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAs 834  
834 ---GAGCTACTGCGCATTCGTGATGCTGCATCAACTG-----GAAAAGGA 2297  
834 pProGluProLysTrpPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArg 854  
854 CTACACAGCTGGGATCACTTATATTGTTGGTGCAAGAACGCCATCACACCCGCTTTTC-- 2355  
854 gArgMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspVal 874

2356	-TGTGCTGCACAAGAAATGAGCGAATTGGGAAGAGTGGTAACATCCAGCT	2403
874	lAlaValAlaAlaVallysGlnTTPGluGluAspMetLysGlnSerLysGluThrGlyI	894
2403	-----	2403
894	eValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAs	914
2404	-----GGGACCACAGTGGACACCAACATCACCACCACCATTTGAGTTTGA	2447
914	pPhePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrVa	934
2448	CTTCTATCTGTGCACCCAGCAGCATCCAGGGCCACGACCATCCCATTA	2507
934	lMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAl	954
2508	TCTTTGGATGACAAACCGTTTCACAGCAGATGAGCTCCAGATCCTGAC	2567
954	aPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHi	974
2568	CCACATTAGTAGTAGCAGCACGCTCTGCTATCCAGACCTGCCCTACTATGCCCCG	2627
974	sLeuSerCysGlnLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAs	994
2628	CCTGGTGGCTTTCGGGGCAGCA---TACCACCTGGTGCAAGAGGAGCATGACAGTGGAGA	2684
994	p 994	
2685	G 2685	

## RESULT 8

US-11-035-669-3

; Sequence 3, Application US/11035669

: Publication No. US20050182015A1

: GENERAL INFORMATION:

APPLICANT: Donna T. Ward

APPLICANT: DONNA I. WATKINS  
APPLICANT: ANDREW T. WATKINS

APPLICANT: Andrew I. Wall  
TITLE OF INVENTION: ANTISENSE MODULATION OF EIF2C1 EXPRESSION

; TITLE OF INVENTION: ANTISENSE  
: FILE REFERENCE: ANDI-001115 P1

```

; FILE REFERENCE: BNDL-VOL1US.PI
: CURRENT APPLICATION NUMBER: US/11/025 669

```

; CURRENT APPLICATION NUMBER: US/1  
 : CURRENT FILING DATE: 2005-01-14

; CURRENT FILING DATE: 2005-01-14  
 : PRIOR PUBLICATION NUMBER: 09,793,907

;; PRIOR APPLICATION NUMBER: 09/

; PRIOR FILING DATE: 2001-02-23  
: PRIOR APPLICATION NUMBER: 08/

PRIOR APPLICATION NUMBER: 08-0679

;; PRIOR FILING DATE: 2001-09-13

;  
PRIOR APPLICATION NUMBER: 09/954

; PRIOR FILING DATE: 2001-09-12

;  
PRIOR APPLICATION NUMBER

**; PRIOR FILING DATE: 2001-11-15**

; NUMBER OF SEQ

; SEQ ID NO 3

; LENGTH: 74

; TYPE: DNA

**; ORGANISM: Homo sapiens**

**FEATURE:**

**FEATURE:**

; NAME/KEY: CDS

LOCATION: (

**Alignment Scores:**

Assignment Scores:  
Pred. No.: 1.13e-41

```
FREQ: NO.: 1.13E-41
Score: 577.00
```

score:	377.00	Macches
Percent similarity:	40.28	Congery

Percent Similarity: 40.2%

Best Local Similarity: 23.2%

Query Match: 10.8%  
Indels: 13  
Canc: 13

DB: 13  
Gap: 13

Year	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050	2055	2060	2065	2070	2075	2080	2085	2090	2095	2100	
Population (millions)	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.0	
GDP (trillion USD)	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5	
Life expectancy (years)	45	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	140	145	150	155	160	165	170	175	180	185	190

US-10-645-746-3 (1-1020) x US-11-035-669-3 (1-

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

Qy 21 MetLysTrpLeuAlaArgProThrGlyLysCysAspG

[illegible][illegible]









Db	570	-----	570	Db	1417	CAAGATACACCAAAAGGCTCAATGAGAAGCAAGTACTGCTATTGCTAAAAGCTACTGCTGC	1476
Qy	118	ThrValTyrSerValCysArgLeuAenThrValThrSerLysMetLeuValSerGluLys	137	Qy	443	ArgLysProHisAspTyrLysGluAenThrLeuLysMetLeuLysGluLeuAspPheSer	462
Db	571	-----	571	Db	1477	CAACGACCCCTGAT---AGAGAGAACTCGATCAAAAACCTTGGTTGTGAAAAAATAATTAC	1533
Qy	138	ValValLysLysAspSerGlu-----LysLysAspGluLysAspLeuGluLysLys	154	Qy	463	SerGluGluLeuAenPheValGluArgPheGlyLeuCysSerLysLysLeuGlnMetIleGlu	482
Db	610	-----	610	Db	1534	AATGATGATCTGAGC-----AAGGAGTTTGGATGTCAGTACTACCACTAGCTGCTCG	1587
Qy	155	IleLeuTyrThrMet-----IleLeuThrTyrArgLysLysPheHisLeuAsnPheSer	172	Qy	483	CysProGlyLysValLeuLysGluProMetLeuValAenSerValAenGluGlnIleLys	502
Db	661	AGCCTCTATACCTGCTGCTCCATTACCTTTTGACTCGAAAGAGTTGTTGTGTAATCTGGCG	720	Db	1588	ATTGAAGCTCGTGTATTTCCCCACCGGATGTTGAAGTACCATGACAGTGGTAAAGAGAAA	1647
Qy	173	ArgGluAenProGluLysAspGluGluAaenArgSerTyrLysPhe-----LeuLysAen	191	Qy	503	MetThrProValIleArgGlyPheGlnGluLysGlnLeuAenValValProGluLysGlu	522
Db	721	GAGAAAAGAGCTGACGGTTCTCTCGGAAGGACAGACCGTTTAAAGTTGCTGTAAGAAT	780	Db	1648	ATGTAAATCCAAAGGCTAGGA-----CAGTGAACATGATGATGACAGAAA-----	1692
Qy	192	Val-----	200	Qy	523	LeuCysCysAlaValPheValAaenGluThrAlaGlyAenProCysLeuGluGluAaen	542
Db	781	GTGACACGACGATCTTTATACAGTTGCAACAGTTCTTGTGCTGAAGCAAGAGGCT	840	Db	1692	-----	1692
Qy	201	ProPheValAenGluGluLysValGlnPheAlaLysAenPheValTyrAspAenAen	220	Qy	543	AspValValLysPheTyrThrGluLeuIleGlyCysLysPheArgGlyIleArgIle	562
Db	841	CCATAT-----	894	Db	1693	-----	1719
Qy	221	SerIleLeuArgValProGluSerPheHisAspProAenArgPheGluGlnSerLeu---	239	Qy	563	GlyAlaAenGluAenArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAenGluTyr	582
Db	895	GATTATGCTCTGTTGGGAGGCTCTTT-----	939	Db	1720	TGGACTTCGGAATTTAAGCCTCAACCTGCTATTCGTTTCATCTCTTGCCCCCTGAACAT	1779
Qy	240	-----	251	Qy	583	AlaPheTyrLysAenCysThrLeuAenThrGlyIleGlyArgPheGluIleAlaIleThr	602
Db	940	AAGGACGCAAGAGATGTTAGGCGTGCCTGGAGATGGTATTGAGTACTGGAGAGTTAT	999	Db	1780	-----	1797
Qy	252	TyrIleGlyIleLysGluLeuPheAspGlyGluProValLeuAenPheAlaIleValAsp	271	Qy	603	GluAlaLysAenMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPheIle	622
Db	1000	TTCCAAAGCTAAAG--CTGACTCAGATGGGTGTTGCTCTCGAACATGACGTTTCAGCA	1056	Db	1798	GAT-----	1836
Qy	272	LysLeuPheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuIleValAsp	291	Qy	623	IleIleSerLysArgGlnLeuAenAlaTyrGlyPheValLysHisTyrCysAspHisThr	642
Db	1057	AGATCATTTTAT-----	1071	Db	1837	GTAATATTGCCGTGATGCTGCTGATCATATGGAATAATAAAGATCTGTGAAACAGAA	1896
Qy	292	ProGlnSerCysAenAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLys	311	Qy	643	IleGlyValAlaAenGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeu	662
Db	1072	CCGATTGTTGTCACGCTTTATTAGCAAGTTTCTGAAATATAGGACTTA-----	1125	Db	1897	TTGGGGATTGCTCTCTCAGTGTGCCAACCTAGACAAAGTTAATAAATCTC-----	1944
Qy	312	MetThrIleArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAenLeuLysLeu	331	Qy	663	ArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAenAlaLys	682
Db	1126	AGGCCACTTAGAGACTCAGATCGACTTAAGGTGAAGAAAGTTTTCAGGACACTGAAAGTT	1185	Db	1945	-----	1989
Qy	332	LysCysAlaGluValTyrAspAenGluMetSerArg-----	343	Qy	683	LeuGlyGlyIleAenGlnGluLeuAspTyrSerGluIleAlaGluIleSerProGluGlu	702
Db	1186	AAAGTTGCTTCAC---TGGAAACGGCACAAAAGTGCACAAATTTAGTGGGATTTCTAGTCTA	1242	Db	1990	ACTGGGGAAGGAACACTGTTCTTAAT-----	2016
Qy	344	---LeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAenSerLeuValTyr	362	Qy	703	LysGluArgLysThrMetProLeu-----	717
Db	1243	CCCATCAGGAGCTAAGTTTCTCTGGAGGCAAAATCAGAGACGCTGTTCAATAT	1302	Db	2017	GATGCTATTAGAAGAAACATACCTCTTACTGATCGTCCAAACCATCATCATGGGTGCT	2076
Qy	363	LysValThrGlyLysSerAspArgGlyArgAenAlaLysLysLysAspThrThrLeuPhe	382	Qy	718	AspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaIleValAlaIle	737
Db	1303	-----	1320	Db	2077	GATGTGACTACCCACAGCCTGGAGAGACTCAAGTCTCTTCTTCTGCTGCTGTGGGCC	2136
Qy	383	LysIleTyrGluGluAenLysLysPheIleGluPheProHisLeuProLeuValLysVal	402	Qy	738	SerIleAen---ProGlyGlyThrIleTyrArgAenMetIleValThrGln-----	753
Db	1321	-----	1359	Db	2137	TCTATGGAGCTGGCCTGAGATAAACAATACCGAGATGGTTGCTTCTGCTCAGCTCATAGG	2196
Qy	403	LysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysPro	422	Qy	754	GluGlu-----	765
Db	1360	GGGAGTGACAAAGCCGCTCTACCTACCAATGGAGCTCTGCCAAATTCACGAA---GGG	1416	Db	2197	GAAGAAATTTATTCAGGACCTGTATAAGCTGGTTTCAGGATCCACAACTGGGTAGTCCAC	2256
Qy	423	GlnArgTyrLysAenArgIleAspLeuValMetGlnAenAspLysPheLeuLysArgAlaThr	442	Qy	766	GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla	785
				Db	2257	-----	2298





QY 221 SerIleLeuArgValProGluSerPheHisAaspProAsnArgPheGluGlnSerLeu--- 239  
Db 895 GATTATGTCCTCTGTTGGGAGGCTCTTT-----TTCCACACTAGTTTGGGA 939  
QY 240 -----GluValAlaProArgIleGluAlaTrpPheGlyIle 251  
Db 940 AAGGACGCAAGAGATGGTAGGGGTGAGCTTGGAGATGGTATTGAGTACTGGAGAGGTTAT 999  
QY 252 TyrIleGlyIleLysGluLeuPheAaspGlyGluProValLeuAsnPheAlaIleValAasp 271  
Db 1000 TTCCAAAGCTTAAGG---CTGACTCAGATGGGTTTGTCTCTGCAACATTGACGTTTCAGCA 1056  
QY 272 LysLeuPhePheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAasp 291  
Db 1057 AGATCATTTTAT-----GAA 1071  
QY 292 ProGlnSerCysAsnAaspValArgLysAaspLeuLysThrLysLeuMetAlaGlyLys 311  
Db 1072 CCGATTGTTGTCACCTGACTTATTAGCAAGTTTCTGAATATAAGGACTTA-----AAC 1125  
QY 312 MetThrIleArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeu 331  
Db 1126 AGGCCACTTAGAGACTCAGATCGACTTAAGGTGAAGAAAGTTTTCAGGACACTGAAAGTT 1185  
QY 332 LysCysAlaGluValTrpAaspGluMetSerArg----- 343  
Db 1186 AAGTGCCTTCAC---TGGAAACGGCACAAGGTCGCAAAATAGTGGGATTTCTAGTCTA 1242  
QY 344 ---LeuThrGluArgHisLeuThrPheLeuAaspLeuCysGluGluAsnSerLeuValTyr 362  
Db 1243 CCCATCAGGAGCTAAGGTTCACTCTGGAGGACAAATCAGAGAAGCGTTGTCAATAT 1302  
QY 363 LysValThrGlyLysSerAaspArgGlyArgAsnAlaLysIleLysTyrAaspThrThrLeuPhe 382  
Db 1303 -----TTTGCTGAAATAATAATAAT----- 1320  
QY 383 LysIleTyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysVal 402  
Db 1321 -----TATAGA-----GTGAATACCAAGCTCTACCTGCTATTCANACA 1359  
QY 403 LysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysPro 422  
Db 1360 GGGAGTGACACAAGACCCGCTTACCTACCCTGAGGAGCTCTGCCAATTCACGAA---GGG 1416  
QY 423 GlnArgTyrLysAsnArgIleAaspLeuValMetGlnAaspLysPheLeuLysArgAlaThr 442  
Db 1417 CAAGAATACACCAAAAGGCTCAATGAGAAGCAAGTGACTGCATTTGCTAAAGACTACCTGC 1476  
QY 443 ArgLysProHisAaspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAaspPheSer 462  
Db 1477 CAACGACCCCTGAT---AGAGAACTCGATCAAAAACCTTGGTTGTGAAAATAATTAC 1533  
QY 463 SerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGlu 482  
Db 1534 AATGATGATCTGAC-----AAGGAGTTTGGGATGTCAGTGACTACCCAACTAGCCTCG 1587  
QY 483 CysProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLys 502  
Db 1588 ATTTAAGAGCTCGTGACTTCCCCACCGATGTTGAAGTACCATGACAGAGTGTTAAAGAGAAA 1647  
QY 503 MetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValProGluLysGlu 522  
Db 1648 ATGGTAATCCAAAGGCTAGA-----CAGTGGACATGATTGACAGAAA--- 1692  
QY 523 LeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsn 542  
Db 1692 ----- 1692  
QY 543 AspValValLysPheTyrThrGluLeuIleGlyGlyCysLysPheArgGlyIleArgIle 562  
Db 1693 -----ATGGTTAATGGAGCAAAA-----GTCACCTCT 1719

QY 563 GlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAaspAlaThrLysAsnGluTyr 582  
Db 1720 TGGACTTGGGAATTTAAGCTCAACTGCTATTTCGGTTCATCTCTGTGTCCTCCCTGAACAT 1779  
QY 583 AlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAlaAlaThr 602  
Db 1780 -----ATTGAGGAAGCTCTCTC 1797  
QY 603 GluAlaLysAsnMetPheGluArgLeuProAaspLysGluGlnLysValLeuMetPheIle 622  
Db 1798 GAT-----ATCCACAAAAGGCGACCTGTCTCCAA-----CTGTGATTT 1836  
QY 623 IleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAaspHisThr 642  
Db 1837 GTAATATTGCCGTGATGTGACTGATCATATGAAAAATAAAGGATCTGTGAAACAGAA 1896  
QY 643 IleGlyValAlaAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeu 662  
Db 1897 TTGGGGATTGCTCTCTCAGTGTGCCAACCTAGACAAGTTAATAAACTC----- 1944  
QY 663 ArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsnAlaLys 682  
Db 1945 -----AACAGCAGTACATGGAATAATGTTGCCCTTGAAGATCAATGTCAG 1989  
QY 683 LeuGlyGlyIleAsnGlnLeuAaspTrpSerGluIleAlaGluIleSerProGluGlu 702  
Db 1990 ACTGGGGNAGGAACACTGTTCTTAAT----- 2016  
QY 703 LysGluArgArgLysThrMetProLeu-----ThrMetTyrValGlyIle 717  
Db 2017 GATGCTATTAGAAGAACAATACCTCTTATTACTGATCGTCCAACCATCATCATGGGTGCT 2076  
QY 718 AspValThrHisProThrSerTyrSerGlyIleAaspTyrSerIleAlaValAla 737  
Db 2077 GATGTCATCCACCCAGCCTGGAGAGACTCAAGTCTTCTATTGCTGCTGTGTGGCC 2136  
QY 738 SerIleAsn---ProGlyGlyThrIleTyrArgAsnMetIleValThrGln----- 753  
Db 2137 TCTATGGACTGCGCTGAGATAACAATAACCGAGGATTGGTTTCTGCTCAAGCTCATAGG 2196  
QY 754 GluGlu-----CysArgProGlyGluArgAlaValAlaHis 765  
Db 2197 GAAGAAATTTATTCAGGACCTGTATAAGCTGGTTTCAGGATCCACACGTCGGTGTGCCAC 2256  
QY 766 GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla 785  
Db 2257 -----TCTGGTTTGATAAGGACATTTTCATAGCATTCAGGAGAGCT----- 2298  
QY 786 GluAsnAsnAaspAsnArgAlaProAlaHisIleValValTyrArgAaspGlyValSerAasp 805  
Db 2299 -----ACAGGCCAGATACCTCAAGGATCATCTTATCGTGACGAGTAAGCGAA 2349  
QY 806 SerGluMetLeuArgValSerHisAaspGluLeuArgSerLeuLysSerGluValLysGln 825  
Db 2350 GGGCAGATTAGTCAGGTTCTGCTACATGAGATGACTGCTATCCGCAAGCCTTGTAACCT 2409  
QY 826 PheMetSerGluArgAaspGlyGluAaspProGluProLysTyrThrPheIleValIleGln 845  
Db 2410 CTCCAA-----GAGAAATATGTTCTCTCGTGTACTTTCTGTGATGTGCCAG 2454  
QY 846 LysArgHisAsnThrArgLeuLeuArgMetGluLysAaspLysProValValAsnLys 865  
Db 2455 AAACGTCACACACACGTTTGTTCCTGACCAACACGGG-----AATCGT 2499  
QY 866 AspLeuThrProAlaGluThrAaspValAlaValAlaLysGlnTrpGluGluAasp 885  
Db 2500 GATATGACT----- 2508  
QY 886 MetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAaspLys 905  
Db 2509 -----GATAAGAGTGGCAATATTCAACCA-----GGTACTCGTGGGACACT 2550  
QY 906 LeuIleValSerLysTyrLysPheAaspPheLeuAlaSerHisGlyValLeuGly 925

```
Db 2551 AAAATCGTCACCTTAATGAAATTCGACTTCTATTGTAACAGCCATGCTGTATTCAGGGA 2610
Qy 926 ThrSerArgProGlyHisThrValMetTyrAspAspLysGlyMetSerGlnAspGlu 945
Db 2611 ACAAGCAGGCGCGCACATTACCATGTACTTCTCGATGAGAAGCGTTTCACCGCTGATCAG 2670
Qy 946 ValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProlIleSer 965
Db 2671 TTGCAATGCTCAACAACACCTCTGCTACAGCTATGCGAGGTGTACAAAATCTGTGTCA 2730
Qy 966 LeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArg 985
Db 2731 ATTGTCCACCAACCTACTACGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2778
Qy 986 ThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThrArg 1001
Db 2779 TACTACATGAGAGTGCAGATGCTCGATGGAGGTTCGAGCAGGTCCTCAGG 2826
```

## RESULT 12

US-10-412-699B-751

; Sequence 751, Application US/10412699B

; Publication No. US20040045049A1

; GENERAL INFORMATION:

; APPLICANT: Mendel Biotechnology, Inc.

; APPLICANT: Zhang, James

; APPLICANT: Fromm, Michael E.

; APPLICANT: Heard, Jacqueline E.

; APPLICANT: Riechmann, Jose Luis

; APPLICANT: Adam, Luc J.

; APPLICANT: Broun, Pierre E.

; APPLICANT: Pineda, Omaira

; APPLICANT: Reuber, T. Lynne

; APPLICANT: Keddle, James S.

; APPLICANT: Yu, Guo-Liang

; APPLICANT: Jiang, Cai-Zhong

; APPLICANT: Samaha, Raymond R.

; APPLICANT: Pilgrim, Marsha L.

; APPLICANT: Creelman, Robert A.

; APPLICANT: DuBell, Arnold N.

; APPLICANT: Ratcliffe, Oliver

; APPLICANT: Kumimoto, Roderick

; APPLICANT: Sherman, Bradley K.

; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants

; FILE REFERENCE: MBI-0048CIP

; CURRENT APPLICATION NUMBER: US/10/412,699B

; PRIOR FILING DATE: 2003-04-10

; PRIOR APPLICATION NUMBER: 09/394,519

; PRIOR FILING DATE: 1999-09-13

; PRIOR APPLICATION NUMBER: 09/489,376

; PRIOR FILING DATE: 2000-01-21

; PRIOR APPLICATION NUMBER: 09/506,720

; PRIOR FILING DATE: 2000-02-17

; PRIOR APPLICATION NUMBER: 09/533,030

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/533,392

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/533,029

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/532,591

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/533,648

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/713,994

; PRIOR FILING DATE: 2000-11-16

; PRIOR APPLICATION NUMBER: 09/819,142

; PRIOR FILING DATE: 2001-03-27

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 311

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 751

; LENGTH: 2910

; TYPE: DNA

```
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1149
US-10-412-699B-751
Alignment Scores:
Pred. No.: 4e-42 Length: 2910
Score: 575.50 Matches: 237
Percent Similarity: 39.0% Conservative: 159
Best Local Similarity: 23.3% Mismatches: 337
Query Match: 10.8% Indels: 283
DB: 8 Gaps: 41
US-10-645-746-3 (1-1020) x US-10-412-699B-751 (1-2910)
Qy 38 LysLysValLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGlu 57
Db 472 AGAAGAGTCATGCTGCTGCGAAT-----CATTTCTTGTTCAGTTGCTGATCGTAT 525
Qy 58 TyrTyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysPro 77
Db 526 CTCACCATACGATGTTTCGATCAATCTCGAGGTTATATCAAAG----- 570
Qy 78 PheProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeu 97
Db 570 ----- 570
Qy 98 ArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAsp 117
Db 570 ----- 570
Qy 118 ThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLys 137
Db 571 -----ACAGTGAACAGAAACGTCATGAACCTCTCGTTAAGAAAT--- 609
Qy 138 ValValLysLysAspSerGlu-----LysLysAspGluLysAspLeuGluLysLys 154
Db 610 -----TATAAGACTCTCATTGGGAGGAAAGTACCAGCGTATGAT---GGAAGGAAA 660
Qy 155 IleLeuTyrThrMet-----IleLeuThrTyrArgLysLysPheHisLeuAsnPheSer 172
Db 661 AGCCCTATACTGCTGCTGCTCCATTACCTTTTGACTCGAAGAGTTTGTGTGAATCTGGCG 720
Qy 173 ArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyrLysPhe---LeuLysAsn 191
Db 721 GAGAAAGAGCTGACGGTTCCTCTGGAAGGACAGACCGTTTAAAGTTGCTGTAAGAAT 780
Qy 192 Val-----MetThrGlnLysValArgTyrAla 200
Db 781 GTGACAAAGCACTGATCTTTATCAGTTGCAACAGTTCCTTGATCGTAAGCAAGAGAGGCT 840
Qy 201 ProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyrAspAsnAsn 220
Db 841 CCATAT-----GATACTATCCAGTCTTGATGTTGTCTTAGGATAAGCCCTCTAAT 894
Qy 221 SerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGlnSerLeu--- 239
Db 895 GATTATGCTCTGTTGGGAGGTCCTTT-----TTCCACACTAGTTTGGGA 939
Qy 240 -----GluValAlaProArgIleGluAlaTrpPheGlyIle 251
Db 940 AAGGACGCAAGAGATGTTAGGGGTGAGCTTGGAGATGGTATTGAGTACTGGAGAGGTAT 999
Qy 252 TyrIleGlyIleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAsp 271
Db 1000 TTCCAAAGCTAAGG---CTGACTCAGATGGGTTGTCTCTGACACATTCAGCTTCAGCA 1056
Qy 272 LysLeuPheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuValAsp 291
Db 1057 AGATCATTTTAT-----GAA 1071
Qy 292 ProGlnSerCysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLys 311
```

Db 1072 CCGATTGTTGTCACCTGACCTTTATTAGCAAGTTTCTGAATATAAGGGACTTA-----AAC 1125  
QY 312 MetThrIleArgGlnAlaalaArgProArgIleArgGlnLeuLeuGlnLeuLeuVal 331  
Db 1126 AGGCCACTTAGAGACTCAGATCGACTTAAGGTGAAGAAAGTTTGTGAGGACACTGAAGATT 1185  
QY 332 LysCysAlaGluValTrpAspGluMetSerArg----- 343  
Db 1186 AAGTGTCTTCAC---TGGAAACGGGCACAAAAGTGCACAAATATAGTGGGATTTCTAGTCTA 1242  
QY 344 ---LeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGluValSerLeuValTyr 362  
Db 1243 CCATCAGGAGCTAAGGTTCACTCTGGAGGACAAATCAGAGAGCGGTGTCTCAATAT 1302  
QY 363 LysValThrGlyLysSerAspArgGlyArgAsnAlaLysGlyTyrAspThrThrLeuPhe 382  
Db 1303 -----TTTGCTGAAAAATAAAT----- 1320  
QY 383 LysIleTyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysVal 402  
Db 1321 -----TATAGA-----GTGAATACCGGCTCTACCTGCTATTCAACA 1359  
QY 403 LysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysPro 422  
Db 1360 GGGAGTGACACAAGACCCGCTACCTACCAATGGAGCTCTGCCAATTCACGAA---GGG 1416  
QY 423 GlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr 442  
Db 1417 CAAGATACACCAAAAGGCTCAATGAGAAAGCAAGTGCATTTGTGTAAGGCTACCTGC 1476  
QY 443 ArgLysProHisAspTyrLysGluLeuThrLeuLysMetLeuLysGluLeuAspPheSer 462  
Db 1477 CAACGACCCCTGAT---AGAGAGAACTCGATCAAAAACCTTGTGTAAGGCTCAATAT 1533  
QY 463 SerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGlu 482  
Db 1534 AATGATGATCTGAGC-----AAGGAGTTTGGGATGTGAGTACCTACCCAACTAGCCTCG 1587  
QY 483 CysProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLys 502  
Db 1588 ATTGAAGCTCGTGTACTCCCCACCGGATGTTGAAGTACCATGACAGTGGTAAAGAGAAA 1647  
QY 503 MetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGluLysGlu 522  
Db 1648 ATGGTAAATCCAAAGCTAGGA-----CAGTGAACATGATTCACAGAAA--- 1692  
QY 523 LeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsn 542  
Db 1692 ----- 1692  
QY 543 AspValValLysPheTyrThrGluLeuIleGlyGlyCysLysPheArgGlyIleArgIle 562  
Db 1693 -----ATGGTTAATGGAGCAAAA-----GTCACTTCT 1719  
QY 563 GlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAsnGluTyr 582  
Db 1720 TGGACTTGGAAATTTAAGCTCACTCGTATTCCGTTTCTCTTGTCCCTGACAT 1779  
QY 583 AlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAlaAlaThr 602  
Db 1780 -----ATTGAGGAAGCTCTTCTC 1797  
QY 603 GluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPheIle 622  
Db 1798 GAT-----ATCCAAAAGGGCACCTGGTCTCCAA-----CTGTTGATT 1836  
QY 623 IleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAspHisThr 642  
Db 1837 GTAATTCCTGATGTGACTGATCATATGGAAAAATTAAGAGTCTGTGAACAGAA 1896  
QY 643 IleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAspLeu 662  
Db 1897 TTGGGGATTGTCTCTCAGTGTGTCACCACTAGACAAGTTAATAAAATC----- 1944

QY 663 ArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsnAlaLys 682  
Db 1945 -----AACAAAGCAGTACATGGAAATATGTTGCTTGAAGATCAATGCAAG 1989  
QY 683 LeuGlyGlyIleAsnGlnGluLeuAspTrpSerGluIleAlaGluIleSerProGluGlu 702  
Db 1990 ACTGGGGAGNAACACTGTTCTTAAT----- 2016  
QY 703 LysGluArgArgLysThrMetProLeu-----ThrMetTyrValGlyIle 717  
Db 2017 GATGCTATTAGAAAGAACATACCTCTTATTACTGATCGTCCAAACCATCATGSGTGCT 2076  
QY 718 AspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValAla 737  
Db 2077 GATGAGTACCCACCCAGCCTGGAGAGGACTCAAGTCTTCTATTGCTGCTGTGTGGCC 2136  
QY 738 SerIleAsn---ProGlyGlyThrIleTyrArgAsnMetIleValThrGln----- 753  
Db 2137 TCTATGGACTGCGCTGAGATAACAATACCGAGGATTGGTTTCTGCTCAAGCTCATAGG 2196  
QY 754 GluGlu-----CysArgProGlyGluArgAlaValAlaHis 765  
Db 2197 GAAGAAATTTATTCAGGACCTGTATAAGCTGGTTTCAGGATCCCAACGCTGGGTAGTCCAC 2256  
QY 766 GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla 785  
Db 2257 -----TCGGTTTGATAAGGGAACATTTTCATAGCATTCAGGAGACT----- 2298  
QY 786 GluAsnAsnAspAsnArgAlaProAlaHisIleValValTyrArgAspGlyValSerAsp 805  
Db 2299 -----ACAGGCCAGATACCTCAAGGATCATCTTATCGTGACGAGTAAGCGAA 2349  
QY 806 SerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGln 825  
Db 2350 GGGCAGTTTAGTCAGGTTCTGCTACATGAGTACGCTGCTATCCGCAAGCTTGTAACTCT 2409  
QY 826 PheMetSerGluArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGln 845  
Db 2410 CTCCAA-----GAGATATTATGTTCTCTCGTTACTTTCGTGATGTCCAG 2454  
QY 846 LysArgHisAsnThrArgLeuLeuArgArgMetGluLysAspLysProValValAsnLys 865  
Db 2455 AAACGTCACACACACGTTTGTCTCCCTGAGCAACACGGG-----AATCGT 2499  
QY 866 AspLeuThrProAlaGluThrAspValAlaValAlaValLysGlnTrpGluGluAsp 885  
Db 2500 GATATGACT----- 2508  
QY 886 MetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLys 905  
Db 2509 -----GATAAGTGGCAATATTCAACCA-----GGTACTGCTGGACACT 2550  
QY 906 LeuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHisHisGlyValLeuGly 925  
Db 2551 AAAATCTGTCCACCTAAATGAAATTCGATCTCTATTGAAACAGCATCTGCTGATAGGGA 2610  
QY 926 ThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGlu 945  
Db 2611 ACAAGCGGCCGCGCACATTACCATGTACTTCTCGATGAGAACGGTTTCCCGCTGATCAG 2670  
QY 946 ValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSer 965  
Db 2671 TTGCAAAATGCTCAACAACACCTCTGCTACAGTATGCGAGGTGTACAAAATCTGTGTC 2730  
QY 966 LeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArg 985  
Db 2731 ATTGTGCCACCGCTACTACGCTCCTTGGCTGCTATTCGGTGCC-----CGC 2778  
QY 986 ThrTyrLysGluHisTyrIleGlyAspTyrAlaGlnProArgThrArg 1001  
Db 2779 TACTACATGAGAGTGAGATGCTGTGATGAGGTTCGAGCGAGGTCCAGG 2826



```
Db 1443 CAGGGCGTCTGGGACATCGGGAACAGCAGATTCCACACGGGCATCGAGATCAAGGTGG 1502
Qy 524 CysCysAlaValPheValValAenGluThrAlaGlyAsnProCysLeuGluAenAsp 543
Db 1503 GCCATTGGTGTTCGCCCCCAG-----CGCCAGTCAGCGGAAGTCCAT--- 1547
Qy 544 ValValLysPheThrThrGluLeuLeu-----GlyGly-CysLysPhe 557
Db 1548 ---CTGAAGTCTTCACAGACAGCTCAGAAAGATCTCGAGAGACGCCGCGCATGCCATC 1604
Qy 557 eArgGlyLeuArgGlyLeuAenGluAenArgGlyAlaGlnSerIle-----MetTy 575
Db 1605 CAGGGCCAGCGGTCTTCGAAATACGCG-CAGGGGGCGGACAGCGGTGGAGCCCATGTT 1663
Qy 575 rAspAlaThrLysAenGluThrAlaPheTyLysAenCysThrLeuAenThrGlyIleG1 595
Db 1664 CCGGCACCTGAAGACACGTATGCG----- 1688
Qy 595 yArgPheGluIleAlaAlaThrGluAlaLysAenMetPheGluArgLeuProAspLysG1 615
Db 1688 ----- 1688
Qy 615 uGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAenAlaTyrgLysPheVa 635
Db 1689 ---GGCCTGACAGCTGGTGTGTATCTCTGCGGCAAGACGCCGCGGTACGCCAGGT 1744
Qy 635 llysHisTyrcysAspHisThrIleGlyValAlaAenGlnHisIleThrSerGluThrVa 655
Db 1745 CAAGCGCGGGAGACACGCTGCTGGGATGCCACGCGAGTGGTGCAGATCAAGAACGT 1804
Qy 655 lThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrglnI1 675
Db 1805 G-----CAGAGGACACCGCCACAGACCCCTGTCCAACCT 1837
Qy 675 eAlaLeuLysIleAenAlaLysLeuGlyGlyIleAenGlnGluLeuAspTrpSerGluI1 695
Db 1838 CTGCTGAAGATCAAGCTCAGCTGGAGCGGTGGAGCGGTGAACAATCTCTG----- 1883
Qy 695 eAlaGluIleSerProGluGluLysGluArgArgLysThrMetProLeuThrMetTyrgVa 715
Db 1884 -----CTGCCCGGCGAGCGCGCGCGGTGTTCACAGACCGCGTC---ATCTTTCT 1930
Qy 715 lGlyLeuAspValThrHisProThrSerTyrcysGlyIleAspTyrcysIleAlaAlaVa 735
Db 1931 GCGAGCAGACGTCACCTACCCCGCGCGGATGGGAAGAGCCCTCCATTGCCGCGCT 1990
Qy 735 lValAlaSerIleAenProGlyGlyThrIleTyArgAenMetIleValThrGlnGluG1 755
Db 1991 GTTGGCAGCATGGAGCGCCACCCCAATCGCTACTCGGCCACCGTGGTGGCAGCAG--- 2048
Qy 755 uCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAl 775
Db 2049 -----CACCAGGAGAGATCATCAAGACCTGGCCGC 2080
Qy 775 aLysPheValLysLeuLeuArgGluPheAlaGluAenAenAspAenArgAlaProAlaHi 795
Db 2081 CATGTCGCCGAGCTCTCATCGATTCTACAAGTCCACGCGCTTCAAG---CCACCCCG 2137
Qy 795 sIleValValTyArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspG1 815
Db 2138 CATCATCTTCTACCGGACGGGTCTCTGAAGCCAGTTCACAGCAGGTTCTCCACCACGA 2197
Qy 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835
Db 2198 GTTGTGGCCATCCGTGAGCGCTGTATCAAGCTA-----GAAAAAGACTA 2242
Qy 835 oGluProLysTyThrPheIleValIleGlnLysArgHisAenThrArgLeuLeuArgAr 855
Db 2243 CCAGCCCGGGATCACCTTCATCGTGGTCAGAGAGGACCACACCCGGCTCTTC---TG 2299
Qy 855 gMetGluLysAspLysProValValAenLysAspLeuThrProAlaGluThrAspValAl 875
```



LOCATION: 3528  
OTHER INFORMATION: a, t, c, g, or other  
US-10-467-397-17

## Alignment Scores:

Pred. No.: 4.1e-41 Length: 3580  
Score: 566.00 Matches: 239  
Percent Similarity: 40.0% Conservative: 159  
Best Local Similarity: 24.0% Mismatches: 385  
Query Match: 10.6% Indels: 214  
DB: 8 Gaps: 38

US-10-645-746-3 (1-1020) x US-10-467-397-17 (1-3580)

```

Qy 19 ProGluMetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLys 38
Db 95 CCCATCCAAAGATATGCTTCAAGCTCCACCTAGACCCGACTTGGGACCTCCGGGAGA 154

Qy 39 LysValLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyr 58
Db 155 ACATCAATCAATACGGCCAAATTCCTCGAATGGACATCCCAAAATGGACATCATCAT 214

Qy 59 TyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPhe 78
Db 215 TATGAATTGGATATCAAGCCAGAGAG-----TGC 244

Qy 79 ProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArg 98
Db 245 CCGAGGAGATTAAACAGGGAATCGTGTGAACACATG-----GTCCAG 286

Qy 99 HisGluLysLysGlnThrAspPheIleLeuGluAspTyr-----ValPheAspGluLys 116
Db 287 CACTTTAAACACAG-----ATCTTTGGGATCGGAAGCCGCTTGAACGGCAGG 337

Qy 117 AspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGlu 136
Db 338 AAGAAATCTATACACAGCCATGCCCTTCCGATGGGAGGCAAGGTG-----GAG 398

Qy 137 LysValValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeu 156
Db 389 CTGGAGTCAAGCTGCCAGGAGAGGCAAGGATCGC-----ATCTTCAAGGTGTCCATC 442

Qy 157 TyrThrMetIleLeuThrTyrArgLysPhePheHisLeuAsnPheSerArgGluAsnPro 176
Db 443 AAGTGGGTGCTCGGTGAGCTTGCGAGCGTTTACAGATGCATTTTCAGGCGCGCTGCC 502

Qy 177 GluLysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGln-- 195
Db 503 AGCGTCCCTTTTCAG-----ACGATCCAGGCCCTGGACGTGTGTCATGAGCCTTG 553

Qy 196 ---LysValArgTyrAlaProPheValAsnGluLysValGlnPheAlaLysAsn 214
Db 554 CCATCCATGAGGTACACCCCGTG----- 577

Qy 215 PheValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArg 234
Db 578 -----GGCCGCTCTTCTTCCACCGCTCCGAAGGCTGCTCTAACCT----- 619

Qy 235 PheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGly 254
Db 620 -----CTTGGCGGGGCGGAGAGTGTGGTTTGGCTTCCATCATGTC 661

Qy 255 IleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPhe 274
Db 662 GTCCGCGCTTCTCTC---TGGAAATGATGCTGAATATTGATGTGTGTCAGCAACAGCGTTT 718

Qy 275 TyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSer 294
Db 719 TACAAGGCA-----CAGCCAGTAATCGAGTTTGTGTGAAGTTTGGATTTTAAAGT 772

Qy 295 CysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIle 314
Db 773 ATTGAAGACAA-----CAAAACCTCTG 796

```

```

Qy 315 ArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLys----- 332
Db 797 ACAGATTCCCAAGGGTAAAGTTTACCACAAAGAAATTAAGGCTTAAGAGGTGAGATAACG 856

Qy 333 ---CysAlaGluValTrpAspAsn-----GluMetSerArgLeuThrGluArg 347
Db 857 CACTGTGGGAGAGATGAAGAGAGATACCGGCTCTGCAATGTGACCCCGCGCCCGCCAGT 916

Qy 348 HisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLys 367
Db 917 CACCAACATTC---CGCTGCAGCAGAGCGGCGGAGCGGTGGAGTGCAGC----- 967

Qy 368 SerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGluGlu 387
Db 968 -----GTGGCCAGTAT-----TTCAAG-----GACAGG 991

Qy 388 AsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLys 407
Db 992 CACAAGTTGGTCTTGGCGTACCCCACTCCCATGTTTACAAGTCGACAGGACAGAGAA 1051

Qy 408 GluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsn 427
Db 1052 CACACCTACCTTCCCTGGAGGTCTGTAACTTGTGTCAGACAAAGATGTATTAAAAA 1111

Qy 428 ArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr---ArgLysProHis 446
Db 1112 TTAACGGAC-----AATCAGACCTCAACCATGATGATGATGATGATGATGATGATG 1165

Qy 447 AspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeu 466
Db 1166 GATCGGCAAGAGAGATTAGCAATGATGCGAAGTCAAGTTCATCAACACAGATCCCA 1222

Qy 467 AsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLys 486
Db 1223 ---TAGTCCGTGAATTTGGAATCATGCTCAAGATGAGATGATGATGATGATGATGATG 1279

Qy 487 ValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrProVal 506
Db 1280 GTGCTGCAGCGCCCTCCATCTCTACGGGGCGAGGAATAAAGCTATTATGCGACCCCTG 1339

Qy 507 -----IleArgGlyPheGlnGluLysGlnLeuAsnVal---ValProGluLysGluLeu 523
Db 1340 CAGGGCGCTGCGACATCGCGAAACAAGCATGTTCCACACGGGCGATCGAGATCAAGTGTG 1399

Qy 524 CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp 543
Db 1400 GCCATTGCGTGTCTCGCCCCCAG-----CGCCAGTGCACGGAGTCCAT--- 1444

Qy 544 ValValLysPheTyrThrGluLeuIle-----GlyGly-CysLysPhe 557
Db 1445 ---CTGAAGTCTTTCACAGACAGCTCAGAAAGATCTCGAGAGACGCTGCGATGCCCATC 1501

Qy 557 eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIle-----MetTy 575
Db 1502 CAGGGCCAGCGCTGCTTCTGCAAAATACGCG-CAGGGGGCGGACAGCGGTGGAGCCCATG 1560

Qy 575 rAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleG 595
Db 1561 CCGGCACCTTGAGAACACATGATCG----- 1585

Qy 595 yArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysG 615
Db 1585 ----- 1585

Qy 615 uGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheVa 635
Db 1586 ----GCGCTGCAGCTGGTGGTGTCTCTCGCCCGCAAGACGCCCGGTGACCGCGAGGT 1641

Qy 635 lLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrVa 655
Db 1642 CAAGCGCGTGGGAGACACCGTGTGGGGATGGCCACGACGTGCTGCAGATGAAGACGT 1701

```



```

Qy 655 lThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIrl 675
Db 1702 G-----CAGAGGACCGCCAGCCAGACCCCTGTCCAACCT 1734
Qy 675 eAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTyrSerGluIrl 695
Db 1735 CTGCCTGAAGATCAACGTCAGCTGGAGCGGTGAACACATCTTG----- 1780
Qy 695 eAlaGluIleSerProGluGluLysGluArgArgLysThrMetProLeuThrMetTyrVa 715
Db 1781 -----CTGCCCGAGCGAGCGCGCGGTGTTCACAGACCGCGTC---ATCTTTCT 1827
Qy 715 lGlyIleAspValThrHisGlyThrSerTyrSerGlyIleAspTyrSerIleAlaIaVa 735
Db 1828 GGAGCAGACGTCACATCAACCCCGCGCGGATGGGAAGAAGCCCTCCATTTGCCGCGCT 1887
Qy 735 lValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGl 755
Db 1888 GGTGGCGCAGATGGACCGCCACCCCAATCGCTACTGCGCCACCGTGGCGGTGCAGCAG-- 1945
Qy 755 uCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAl 775
Db 1946 -----CACCGCGCAGGAGATCATACAAGACCTGGCCGC 1977
Qy 775 aLysPheValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAlaHi 795
Db 1978 CATGTGTCGCGAGCTCTCATCCAGTTCTACAGTCCACGCGCTTCAAG---CCACCCGC 2034
Qy 795 sIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspGl 815
Db 2035 CATCATCTTCTACCGCGCGGTGTCTCTGAGCGCAGTTCACGAGTTCTCCACCCACGA 2094
Qy 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835
Db 2095 GTTGTCTGGCCATCCGTGAGCGCTGTATCAAGCTA-----GAAAGACATTA 2139
Qy 835 oGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuArgAr 855
Db 2140 CAGCCCGGGATCACCTTCTATCGTGTGCGAGAAGGACCCACCCCGGCTCTC---TG 2196
Qy 855 gMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAl 875
Db 2197 CACTGACAGACGAGCGGGTTGGGAAAGTGGAAACATTCCAGCA----- 2242
Qy 875 aValAlaAlaValLysGlnThrGluAspMetLysGluSerLysGluThrGlyIleVa 895
Db 2242 ----- 2242
Qy 895 lAanProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPh 915
Db 2243 -----GGCACGACTGTGGACACGAAATCACCCACCCACCGAGTTGCACTT 2289
Qy 915 ePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935
Db 2290 CTACCTGTGTAGTCAGCTGGCATCCAGGGGCAAGCAGCGCTTCGCACTATCACGTCCT 2349
Qy 935 cTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPh 955
Db 2350 CTGGGACGACAAATCGTTTCTCTCTGATGAGTGCAGATCTTAACCTACACGCTGTGCA 2409
Qy 955 eLeuSerAlaAtqCysArgLysProIleSerLeuProValProValHisTyrAlaHisLe 975
Db 2410 CACCTAGTGGCTGCACAGCTCCGTGTTCATCCAGCGCCGACGATACATAGCTCACCT 2469
Qy 975 uSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990
Db 2470 GGTGGCTTCCGGGCCAGG---TACCACCTGGTGGGATAGGAACAT 2512

```

RESULT 15

US-11-097-143-27926  
; Sequence 27926, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:

```

; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27926
; LENGTH: 3486
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-27926

```

Alignment Scores:

```

Pred. No.: 5,448-41 Length: 3486
Score: 564.50 Matches: 233
Percent Similarity: 40.4% Conservative: 174
Best Local Similarity: 23.1% Mismatches: 394
Query Match: 10.6% Indels: 206
DB: 13 Gaps: 37

```

US-10-645-746-3 (1-1020) x US-11-097-143-27926 (1-3486)

```

Qy 38 LysLysValLeuLeuValAenTyrPheLysPheSerSerLysIleTyrAspArgGlu 57
Db 527 CGCCCGATTGTCTGGCGCCCAATCACTTCCAGGTGACA-----ATGCCGCTGGC 577
Qy 58 Tyr---TyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLys 76
Db 578 TATGTGCATCACTATGACATCAATATACAGCCGACAGTGTCCGCGAAAGGTGAACCGT 637
Qy 77 ProPheProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTyrGlnHis 96
Db 638 GAGATT-----ATCGAGACTATGTGTGCATGCTTATAGCAAGATCTTCGA----- 682
Qy 97 LeuArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLys 116
Db 683 -----GTGCTCAAGCCG---GTGTTCCGATGTCGC 709
Qy 117 AspThrValTyrSerValCysArgLeuAenThrValThrSerLysMetLeuValSerGlu 136
Db 710 AACAACTGTATACCCGCGGATCCCTGCCCATTTGGCAACGAGCGTCTA-----GAG 760
Qy 137 LysValValLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeu 156
Db 761 CTGGAGGTTACTCTACCCGCGCGGCGGCAAGATCGA----- 796
Qy 157 TyrThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAenPro 176
Db 797 -----ATCTTTCGCGTGCAGATCAAGTGGCAGGCTCAGGCTCTCGCTCTCAATTG 847
Qy 177 GluLysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLys 196
Db 848 GAGGAAGTCTCTCGAAGGCCGC-----ACGCGGCAG 877

```

197 ValArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheVal 216  
 878 ATACCTATGATGCCATTTTGGCGTCGATGGTGCATCGCCCATCTGCCAGCATGACG 937  
 217 TyrAsp-----AsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArg 234  
 938 TACACGCCAGTGGGAGGTAGCTTCTTCAGTTCCTCCCGAGGGGTACTACCATCCC----- 991  
 235 PheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGly 254  
 992 -----CTGGGTGGTGACGCGAGGTTTGGTTTCGGTTTCCATCAGAGC 1033  
 255 IleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPhe 274  
 1034 GTAAGGCC---TCCGATGGAGATGATGCTCAATATCGATGTTCTGGGCCACCGCTTTC 1090  
 275 TyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSer 294  
 1091 TACAAGGCT-----CAACCAAGTCATTGACTTCATGTGCGAGGTGCTGCACATTCGCGAC 1144  
 295 CysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIle 314  
 1145 ATCAACGAGCAG---CGCAAAACCGCTC----- 1168  
 315 ArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLysCysAla 334  
 1169 ACCGATTCCGACGCGCTCAAGTTCACGAAGGAGATCAAGGGTTTGAAGATCGAGATCACC 1228  
 335 GluValTrpAspAsnGluMetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeu 354  
 1229 CAC----- 1231  
 355 CysGluGluAsnSerLeuValTyrLysValThrGlyLysSerAspArgGly----- 371  
 1232 TCGGCCAGATGCGTCGCAAGTATCTGTGTGCAACGTCATCTGCCGCCCGCTGCAGATG 1291  
 372 -----ArgAsnAlaLysLysTyrAspThrThrLeuPheLysIle 384  
 1292 CAATCATTTCCACTGCAGCTGGAGAACGGACAGACCGTAGAGTGCACCGTGGCCCAAGTAC 1351  
 385 Tyr---GluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLys 403  
 1352 TTCTCTGGCAAGTACCGCATGAAATTGGCTACCCGACCTTGCCTCTGCTGCAGTTGGC 1411  
 404 SerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGln 423  
 1412 CAAGAGCACACACACTTACTTCCTCTAGAGGTGTCAACATT---GTGGCGCGACAG 1468  
 424 ArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArg 443  
 1469 CGGTGCATTAAAGCTGACCGATATGCAGACGTGACCATGATCAAGGCCACAGCTCGT 1528  
 444 LysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSer 463  
 1529 TCAGTCCGGATCGTAGGCGTGAGATTAAACAACCTTGTTAAAGCGCGCGACTTCAACAAC 1588  
 464 GluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCys 483  
 1589 GAT-----TCGTATGTGCAAGAGTTTGGCTCGACCATCTCCAAATTCGATGATGAGGTA 1642  
 484 ProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMet 503  
 1643 CGAGGACGGCTCTTGCTCTCTCCCAAGCTTCAGTATGGGGACGT----- 1687  
 504 ThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGluLysGluLeu 523  
 1688 -----GTGTCTACCGGCTCACCGGCCAGCAGCTG---TTCCCGCCACAGACAAGGTG 1738  
 524 CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp 543  
 1739 AGCTTGCCC-----TCGCCCAACCAAGGTTGATGGGATATCGCA 1777

QY	544	ValVallysPheTyrThrGluLeuLeuIleGlyCysLysPheArgGlyIleArgGlyGly	563
DB	1778	GCACAGCAGTCTCTCACTGGCGTC-----GAGATCCGATCATCGG	1816
QY	564	AlaAsnGluAsnArgGlyAlaGln---SerIleMetTyrAspAlaThrLysAsnGluTyr	582
DB	1817	GCCATCGCCTGTTTGGCCACACGCGCAGCGTGCAGGATCGCTGCCTAAATTCACC	1876
QY	583	AlaPheTyrLysAsnCysThrLeuAsnThrGly-----IleGly-----	595
DB	1877	CAGCAGCTGCAGAGATCTCAACAGATGCAGGCATGCCGATAATTGCACAGCGTGCCTC	1936
QY	596	---ArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLys	614
DB	1937	TGTAAGTACGCCACCGCGCGCGATCAAGTGGAAACCCATGTTCCGTTACCTGGAAGATCACC	1996
QY	615	GluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPhe	634
DB	1997	TTCCCCCGCCTGCACACTGCTCGTGTTGTCTGCCCGGCAAGACTCAGTATACGCCGAG	2056
QY	635	ValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThr	654
DB	2057	GTGAAGCGTGTAGTGTACACCGTCTCGGTATGCGCACCCAGTGTGTGCAGGCCAAGAAC	2116
QY	655	ValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGln	674
DB	2117	GTGAACAAGACA-----TCGCCACAGACGCTCTCTAAT	2149
QY	675	IleAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnLutLeuAspTTPSerGlu	694
DB	2150	CTGTGCTGAAGATCAACGTCNAAGTTGGCGGCATCAATTCAATTCTG-----	2197
QY	695	IleAlaGluIleSerProGluGluLysArgLysThrMetProLeuThrMetTyr	714
DB	2198	GTCCCTCCATTCGGCCNAAGGTTCTCAATGACCG-----GTCATCTTT	2242
QY	715	ValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAla	734
DB	2243	TTGGTCCGATGTGACACACACCCACGAGCTGGCGACACAAAGAAACCATCGATTGCGCC	2302
QY	735	ValValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGlu	754
DB	2303	GTGCTGGGCTCCATGATGCCATCCATCGCGTATGCGCCACCGTTCCGCTACAGCAG	2362
QY	755	GluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGlu	774
DB	2363	-----CACCGACAGAGATCATCCAGGAGCTGAGC	2392
QY	775	AlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAla	794
DB	2393	AGCATGTGTGGCGCGAGCTGTTGATCATGTTCTACAAGTGCAGGGCGGCTACAAGCCCCAC	2452
QY	795	HisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAsp	814
DB	2453	CGCATCATCTCTATCGTGAACGAGTCTCCGAGGACAATTCCACATGCTCTCGCAACAC	2512
QY	815	GluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAsp	834
DB	2513	GAATTGACCCCATTCGCGAGGCGCTGCATTAGCTA-----GAG	2551
QY	835	ProGlu-----ProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeu	852
DB	2552	CCAGAGTATCGCGCGGCATCATCATTTGTTGTGCAGAGCGCCATCACACTGCAGCTC	2611
QY	853	LeuArgArgMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThr	872
DB	2612	TTCTGCGCGAGAGAAGGAG-----	2632
QY	873	AspValAlaValAlaValLysGlnTTPGluGluAspMetLysGluSerLysGluThr	892
DB	2633	-----CAGAGCGCAATCG	2647
QY	893	GlyIleValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLys	912

```

Db      2648 GGC-----ATATTCCTCCGAGCACCACCGTCGATGTGGGCATCACACATCCCACCGAA 2701
Qy      913 PheAspPhePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHisTyr 932
Db      2702 TTTGATTCTCTATGTCAGCCATCAGGGCATCCAGGGCACCAGTCGCCCTCGCACTAC 2761
Qy      933 ThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGly 952
Db      2762 CAGGTTCTGTGGACGACAAATCACTTTGACTCGGACGAGCTGCAGTGCCTCAGTATCAG 2821
Qy      953 LeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyr 972
Db      2822 CTATGCCATACGTACGTGGCTGCACCCGATCCGTCAGTATACCGGGCGCCAGCTACTAC 2881
Qy      973 AlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIle 992
Db      2882 GCCCATTTAGTGGCCTTCGGTGCAGAA--TATCATCTGGTGAGAGAGGAGCAC----- 2932
Qy      993 GlyAsp-TyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnThrAsnVa 1012
Db      2933 ---GATTCGGGGGAGGTTGCGCACGAGCGGCTGCTCAGAGGATCGTACGC---CAGGT 2986
Qy      1012 lLysTyrProGlyMetSer 1018
Db      2987 GCCATGCCCGAGGCCCATCA 3005

```

Search completed: July 6, 2006, 02:32:01  
 Job time : 2800 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 6, 2006, 01:29:24 ; Search time 1437 Seconds  
(without alignments)  
1384.275 Million cell updates/sec

Title: US-10-645-746-3

Perfect score: 5349

Sequence: 1 MSNPFPELEKGFYRHSIDPE.....RHEMHLQTNVVKYQMSFA 1020

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 844069 seqs, 650066433 residues

Total number of hits satisfying chosen parameters: 1688138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/abes/ABSSWEB\_spool/US10645746/runat\_05072006\_110200\_5345/app\_query.fasta.1  
-DB=Published Applications NA\_New -QWMT=fastap -SUPFIX=p2n.rnpbn -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANSHUMAN40.cdi -LIST=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -HOST=abes03p  
-USER=US10645746 @CN 1.1.206 @runat\_05072006\_110200\_5345 -NCPU=6 -ICPU=3  
-NO WMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -XGAPOP=6 -XGAPEXT=7  
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : Published Applications NA\_New:

1: /EMC Celerra SID33/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
2: /EMC Celerra SID33/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
3: /EMC Celerra SID33/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
4: /EMC Celerra SID33/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
5: /EMC Celerra SID33/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
6: /EMC Celerra SID33/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
7: /EMC Celerra SID33/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
8: /EMC Celerra SID33/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
9: /EMC Celerra SID33/ptodata/2/pubpna/US60\_NEW\_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	577	10.8	7478	US-11-266-748A-56189	Sequence 56189, A
2	555	10.4	4758	US-11-266-748A-30266	Sequence 30266, A
3	535	10.0	3474	US-10-953-349-4063	Sequence 4063, Ap
4	510	9.5	3320	US-11-266-748A-31508	Sequence 31508, A
5	520	9.5	3320	US-11-266-748A-56999	Sequence 56999, A
6	509	9.5	3414	US-10-449-902-24843	Sequence 24843, A
7	500	9.3	3432	US-10-449-902-27837	Sequence 27837, A
8	499.5	9.3	3746	US-11-218-305-22486	Sequence 22486, A

9	493	9.2	1793	8	US-11-266-748A-25558	Sequence 25558, A
10	486	9.1	2503	6	US-10-449-902-19320	Sequence 19320, A
11	479	9.0	5408	7	US-11-218-305-543	Sequence 543, App
12	469.5	8.8	3350	7	US-11-218-305-19388	Sequence 19388, A
13	466	8.7	2012	7	US-11-218-305-14683	Sequence 14683, A
14	455	8.5	2723	7	US-11-218-305-19467	Sequence 19467, A
15	455	8.5	3827	7	US-11-218-305-19467	Sequence 19466, A
16	453.5	8.5	2808	8	US-11-216-545-2985	Sequence 2985, Ap
17	443.5	8.3	3178	6	US-10-449-902-14723	Sequence 14723, A
18	443.5	8.1	3003	6	US-10-449-902-8899	Sequence 8899, Ap
19	421.5	7.9	1750	6	US-10-449-902-10173	Sequence 10173, A
20	420.5	7.9	3047	6	US-10-449-902-14194	Sequence 14194, A
21	412.5	7.7	2016	7	US-11-218-305-22484	Sequence 22484, A
22	403.5	7.5	1754	6	US-10-953-349-18472	Sequence 18472, A
23	397	7.4	2001	7	US-11-218-305-22483	Sequence 22483, A
24	396.5	7.4	1816	6	US-10-953-349-19397	Sequence 19397, A
25	392.5	7.3	2095	7	US-11-218-305-15556	Sequence 1556, Ap
26	377	7.0	1710	7	US-11-218-305-19385	Sequence 19385, A
27	374	7.0	3139	7	US-11-218-305-20989	Sequence 20989, A
28	358.5	6.7	1934	8	US-11-216-545-6852	Sequence 6852, Ap
29	344.5	6.4	1211	6	US-10-449-902-26400	Sequence 26400, A
30	340	6.4	1324	8	US-11-216-545-2982	Sequence 2982, Ap
31	331	6.2	1196	6	US-10-449-902-3167	Sequence 3167, Ap
32	331	6.2	1196	6	US-10-449-902-3806	Sequence 3806, Ap
33	331	6.2	1298	8	US-11-216-545-2984	Sequence 2984, Ap
34	312	5.8	3394	8	US-11-266-748A-26618	Sequence 26618, A
35	312	5.8	3394	8	US-11-266-748A-29538	Sequence 29538, A
36	286.5	5.4	1094	7	US-11-218-305-21995	Sequence 21995, A
37	283.5	5.3	965	7	US-11-218-305-22485	Sequence 22485, A
38	260.5	4.9	2456	8	US-11-293-697-1525	Sequence 1525, Ap
39	259	4.8	4753	7	US-11-218-305-19384	Sequence 19384, A
40	258.5	4.8	892	7	US-11-218-305-19386	Sequence 19386, A
41	253.5	4.7	997	8	US-11-216-545-6851	Sequence 6851, Ap
42	249	4.7	2997	6	US-10-449-902-27469	Sequence 27469, A
43	232.5	4.3	772	6	US-10-449-902-4254	Sequence 4254, Ap
44	232	4.3	903	7	US-11-218-305-19468	Sequence 19468, A
45	206	3.9	1000	8	US-11-266-748A-394325	Sequence 394325, A

#### ALIGNMENTS

RESULT 1  
US-11-266-748A-56189  
; Sequence 56189, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; CURRENT FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 483996  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 56189





Db 491 GATAGGTTGATATGAGGTGACTCTTCCAGGCGAGGTAAGACCAACATTTAAAGTG 550  
Qy 155 -----lleLeuTyrThrMetIleLeuThrTyrArgLysLysPheHis 168  
Db 551 TCTGTTTCACTGGGTGTCAGTTGTGTGAGCCCTTCAGTTGCTTTTAAAGCTTTGGCTGGCAC 610  
Qy 169 LeuAenPheSerArgGluAenProGluLysAspGluGluAlaAenArgSerTyrLysPhe 188  
Db 611 TTGAAT-----GAATCCAGATGACTCAGTACAAGCA----- 643  
Qy 189 LeuLysAenValMetThrGln-----LysValArgTyrAlaProPheValAenGlu 205  
Db 644 ---CTTGATGTTATCACAAGACACCTTCCCTCAGGTACACCCAGTG----- 691  
Qy 206 GluIleLysValGlnPheAlaLysAenPheValTyrAspAenAenSerIleLeuArgVal 225  
Db 692 -----GGCCGTTCTCTTTTCTCACCC 712  
Qy 226 ProGluSerPheHisAspProAenArgPheGluGlnSerLeuGluValAlaProArgIle 245  
Db 713 CCGGAAGTTACTACACCT-----CTGGAGGGGCGAGG 748  
Qy 246 GluAlaTrpPheGlyIleTyrIleGlyIleLysGluLeuPheAspGlyGluProValLeu 265  
Db 749 GAGGTCTGTTGTTTTCATCAGTCTGTGAGACCTGCCATG---TGGAATATGATGCTC 805  
Qy 266 AenPheAlaIleValAspLysLeuPheTyrAenAlaProLysMetSerLeuLeuAspTyr 285  
Db 806 AACATTGATGTCTGCAACTGCTTCTACCGGCT-----CAGCCTATCATTTAGTTTC 859  
Qy 286 LeuLeuLeuIleValAspProGlnSerCysAenAspAspValArgLysAspLeuLysThr 305  
Db 860 ATGTGTGAGTTTATGACATTCTAGAACATCATGTAACACACCAA----- 904  
Qy 306 LysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgProArgIleArgGlnLeu 325  
Db 905 -----CCTCTAACACAGCTCCAGCGGTGTCAAAATTACCAAGAA 943  
Qy 326 LeuGluAenLeuLysLeuLys-----CysAlaGluValTrpAspAen----- 339  
Db 944 ATCAGAGGTCTCAAGTTGAGGTGACCCACTGTGGACAGATGAACGAAATACCGAGTT 1003  
Qy 340 ---GluMetSerArgLeuThrGluArgHisLeuThrPhe---LeuAspLeuCysGluGlu 357  
Db 1004 TGTAATGTCACTAGACGGCCAGCAGTCAATCAACTTTCTTCTTGCAGCTA----- 1054  
Qy 358 AenSerLeuValTyrLysValThrGlyLysSerAspArgGlyArgAenAlaLysLysTyr 377  
Db 1055 -----GAAAAACGGTCAAGCTATG 1072  
Qy 378 AspThrThrLeuPheLysIleTyrGluGluAenLysLysPhe---IleGluPheProHis 396  
Db 1073 GAATGTACAGTAGCTCAATATTTTAAAGCAAAAGTATAGTCTGCAACTGAAATACCCCAT 1132  
Qy 397 LeuProLeuLysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeu 416  
Db 1133 CTTCCTGTCTCCAGTGGGACAGAACAAAGCATACATACTGCCACTCGAGTCTGT 1192  
Qy 417 GluValHisGluLysProGlnArgTyrLysAenArgIleAspLeuValMetGlnAspLys 436  
Db 1193 AATATA---GTGGCAGGACGAGTGTATCAAGAGCTCACAGACAATCAGACTTCCACA 1249  
Qy 437 PheLeuLysArgAlaThrArgLysProHisAspTyrLysGluAenThrLeuLysMetLeu 456  
Db 1250 ATGATCAAAAGCTACAGCAAGATCTGCTCTGTGACAGACAGGAAGAGATCAGTAGAGTGGTG 1309  
Qy 457 LysGluLeuAspPheSerSerGluGluLeuAenPheValGlnArgPheGlyLeuCysSer 476  
Db 1310 AAGACCAACAGTATGGTGGGTGACCTGATCCATACCTTAAAGAAATTTGTTATTTGTTGTC 1369  
Qy 477 LysLeuGlnMetIleGluCysProGlyLysValLeuLysGluProMetLeuValAenSer 496  
Db 1370 CACAATGAAATGACAGCTCACAGGAGGTACTTCCAGCAGGCAATGCTGCAATATGGA 1429

Qy 497 ValAenGluGlnIleLysMetThrPro-----ValIleArgGlyPheGln 511  
Db 1430 GGCCTGGAATAAAACAGTAGCCACACCCACAGGGTGTCTGGACATCGCAGGA----- 1483  
Qy 512 GluLysGlnLeuAenVal---ValProGluLysGluLeuCysCysAlaValPheValVal 530  
Db 1484 ---AAGCAGTTTATGCTGGCATTGAAATTAAGTTTGGCAGTTGCTTTTGGCACCT 1540  
Qy 531 AenGluThrAlaGlyAsnProCysLeuGluGluAenAspValValLysPheTyrThrGlu 550  
Db 1541 CAGAAA-----CAATGTAGGAA-----GATTTACTAAAGATTTCACTGAC 1582  
Qy 551 LeuIle-----GlyGlyCysLysPheArgGlyIleArgIleGlyAla 564  
Db 1583 CAGCTCGGTAAATCTTAAGGATCGAGGAATGCCATCCAGGTCAGCCATGTTTCTGC 1642  
Qy 565 AenGluAenArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAenGluTyrAlaPhe 584  
Db 1643 AAGTATGCACAAGGTGCAGACAGTGTG----- 1669  
Qy 585 TyrLysAenCysThrLeuAenThrGlyIleGlyArgPheGluIleAlaThrGluAla 604  
Db 1669 ----- 1669  
Qy 605 LysAenMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPheIleIle 624  
Db 1670 GAGCCTATGTTTAAACATCTGAATAGTATTATGTGGCTACAGTAAATAGTGTATC 1729  
Qy 625 SerLysArgGlnLeuAenAlaTyrGlyPheValLysHisTyrCysAspHisThrIleGly 644  
Db 1730 CTGCTCGAAAGACACACCATATATCGGAGGTGAACGTTGGAGATACCTTCTAGGT 1789  
Qy 645 ValAlaAenGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeuArgHis 664  
Db 1790 ATGGCCACACAGTGTGTCCAGGTAAAAAATGTAGTGAAGACC----- 1831  
Qy 665 GluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAenAlaLysLeuGly 684  
Db 1832 -----TCACCTCAACCTTCCAAATCTTTGCTGAGATTAATGCAAACTTGA 1882  
Qy 685 GlyIleAenGlnGluLeuAspTrpSerGluIleAlaGluIleSerProGluGluLysGlu 704  
Db 1883 GGAATTAACAATGTGCTT-----GTGCTCATCAAGGCC 1918  
Qy 705 ArgArgLysThrMetProLeuThrMetTyrValGlyIleAspValThrHisProThrSer 724  
Db 1919 TCGGTGTTCCAGCAGCTGTC---ATCTTCTCGGAGCGGATGTCCACACCCCCAGCA 1975  
Qy 725 TyrSerGlyIleAspTyrSerIleAlaValAlaSerIleAenProGlyGlyThr 744  
Db 1976 GGGGATGGAGAAACCTTCCATTCCTGCTGTGTGGTGGCAGTATGATGCCACCCAGC 2035  
Qy 745 IleTyrArgAenMetIleValThrGlnGluCysArgProGlyGluArgAlaValAla 764  
Db 2036 CGGTACTGTGCCACCGTTCGGGTGCGGAGCTTCCCGCAGAGAGATCTCCCAAGAGCTCCTC 2095  
Qy 765 HisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPhe 784  
Db 2096 TACAGTCAAGAGGTTCATCCAGACCTGACTAAACATGTTTCGAGAGCTGCTGATTCAGTTC 2155  
Qy 785 AlaGluAenAenAspAenArgAlaProAlaHisIleValValTyrArgAspGlyValSer 804  
Db 2156 TACAATCCACAGCTTCAAA---CCACTCGATCATCTATACCGTGGAGGGTATCT 2212  
Qy 805 AspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLys 824  
Db 2213 GAGGACAAAATGAACAGGTAGCTGGCCAGAACTAATAGCAATTCGA-----AAG 2263  
Qy 825 GlnPheMetSerGluArgAspGlyGluAspProGluProLysTyrThrPheIleValIle 844  
Db 2264 GCATGTATAGTCTGGAA-----GAAGATTACCGGCGAGGAATACTTATATTGTTGGTG 2317



QY 845 GlnLysArgHisAsnThrArgLeuArgMetGluLysAspLysProValValAsn 864  
 DB 2318 CAAAAGACATACACAGACTCTTC---TGTCAGATAAACAGAAAGGTAGGAAA 2374  
 QY 865 LysAspLeuThrProAlaGluThrAspValAlaValAlaValLysGlnTrpGluGlu 884  
 DB 2375 AGTGGCAATGTACCAGCA----- 2392  
 QY 885 AspMetLysGluSerLysGluThrGlyLeuValAsnProSerSerGlyThrValValAsp 904  
 DB 2393 -----GGCACTACAGTGGAT 2407  
 QY 905 LysLeuLeuValSerLysTyrLysPheAspPheLeuAlaSerHisGlyValLeu 924  
 DB 2408 AGTACCATCACACATCATCTGAGTTGACTTTTACTCTGTAGTCATGAGGAATTCAG 2467  
 QY 925 GlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAsp 944  
 DB 2468 GGAACACAGCGTCCCTCACATTACAGGTCTTGTGGGATGACAACTGCTTCACTGCAGAT 2527  
 QY 945 GluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIle 964  
 DB 2528 GAACTCCAGCTACTGACTTACCAGCTGTGTACACCTATATGTAGGTGCACTCGCTCAGTC 2587  
 QY 965 SerLeuProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyr 984  
 DB 2588 TCTATCCAGCCCTGCATATATATGCCGCTTGTAGCATTTAGGGCAAGG---TATCAT 2644  
 QY 985 ArgThrTyrLysGluHis 990  
 DB 2645 CTGGTGGATAAAGATCAT 2662

## RESULT 3

US-10-953-349-4063  
 ; Sequence 4063, Application US/10953349  
 ; Publication No. US20060107345A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ALEXANDROV, Nikolai et al.  
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
 ; TITLE OF INVENTION: ENCODED THERAPY  
 ; FILE REFERENCE: 2750-1579PUS2  
 ; CURRENT APPLICATION NUMBER: US/10/953,349  
 ; CURRENT FILING DATE: 2004-09-30  
 ; NUMBER OF SEQ ID NOS: 40252  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 4063  
 ; LENGTH: 3474  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-10-953-349-4063

## Alignment Scores:

Pred. No.: 2,32e-44 Length: 3474  
 Score: 535.00 Matches: 231  
 Percent Similarity: 38.8% Conservative: 184  
 Best Local Similarity: 21.6% Mismatches: 370  
 Query Match: 10.0% Indels: 284  
 DB: 6 Gaps: 42

US-10-645-746-3 (1-1020) x US-10-953-349-4063 (1-3474)

QY 1 MetSerSerAsnPheProGluLeu-----GluLysGlyPheTyrArgHisSerLeuAsp 18  
 DB 586 CTGGCTCAGCAATTTGAACAACACTCTCTGTTGAACAAGAGAGCTCCCAAGTCAGGCAATCCAG 645  
 QY 19 Pro-----GluMetLysTrpLeuAlaArgProThrGlyLysCysAsp 32  
 DB 646 CCTATACCTCTCTAGCAGGCTTCAAGTTTCCATGAGGCTCGTAAAGACAGAGT 705  
 QY 33 GlyLysPheTyrGluLysLysValLeuLeuValAsnTrpPheLysPheSerSerLys 52  
 DB 706 GGA-----AAGCGTTGCATTGTGAAGGCTAAC-----CATTTCTTTGCTGAA 747

QY 53 IleTyrAspArgGluTyrTyrGluTyrGluValLysMetThrLysGluValLeuAsnArg 72  
 DB 748 CTCCTCATAGGATTTTGACCACTATGATGTTTACCATTACTCCGGAAGTTTACATCAAGG 807  
 QY 73 LysProGlyLysProPheProLysLysThrGluIleProIleProAspArgAlaLysLeu 92  
 DB 808 GGTGTCAATCGTCTGTGATGAACAA----- 834  
 QY 93 PheTrpGlnHisLeuArgHisGluLysGlnThrAspPheIleLeuGluAspTyr--- 111  
 DB 835 -----CTTGTGATAATTATCGT 852  
 QY 112 -----ValPheAspGluLysAspThrValTyrSerVal 122  
 DB 853 GATTCTCACCTTGGAAAGTCGTCTCCAGCGTATATGTCGAAAGAGCTTTTACACTGCT 912  
 QY 123 CysArgLeu-----AsnThrValThrSerLysMetLeuValSerGluLysValValLys 140  
 DB 913 GGTCCACTTCCCTTTAACTCCAAGGAGTTCAGAAATCAATCTTCTTGACGAAAGATGAGG 972  
 QY 141 LysAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeuTyrThrMetIle 160  
 DB 973 GCTGGAGGTCAAAGACGAGAAAGG-----GAATTTAAAGTTGTGATCAAGCTAGTT 1023  
 QY 161 LeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGluLysAsp--- 179  
 DB 1024 ---GCACGTGCTGATCTGCATCACCTAGGAATGTTTTTGGAGGGGAAACAATCAGATGCC 1080  
 QY 180 ---GluGluAlaAsnArgSerTyrLysPhe---LeuLysAsnValMetThrGlnLysVal 197  
 DB 1081 CCACAGAGAGCTCTGCAGGTTCTTGACATTTGTTCTGTCAGCTGCCGACTCTAGATC 1140  
 QY 198 ArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyr 217  
 DB 1141 AGGTATATTCG----- 1152  
 QY 218 AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln 237  
 DB 1153 -----GTGGCCCGTCTCTTTTATTTCCCTCGAT---ATAGGAAAA 1188  
 QY 238 SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyLys--- 256  
 DB 1189 AAACATCATTTGGGGATGGCTTGGAGAGCTGGTGGATTTCTACCAAGCATTCGTCCT 1248  
 QY 257 -----GluLeuPheAspGlyGlu 262  
 DB 1249 ACACAGATGGGCTTATCACTCAATATTGATATGTCATCGACAGCTTCATAGAGCAAC 1308  
 QY 263 ProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAlaProLysMetSerLeu 282  
 DB 1309 CCTGTGATTTCAGTTT-----GTC 1326  
 QY 283 LeuAspTyrLeuLeuLeuIleValAspProGlnSerCysAsnAspAspValArgLysAsp 302  
 DB 1327 TGTGATTCTTAACCGGGATATTTCTCTCGACCTTTATCTGATCTGATCTGTTAAG 1386  
 QY 303 LeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgProArgIle 322  
 DB 1387 ATAAAAAAGGCTCTTAGAGGTGTCAAAGTTGAAGTGAAGTCACTCATCGAGAAACATCGCCGG 1446  
 QY 323 ArgGlnLeuLeuGluAsnLeuLysLysCysAlaGluValTrpAspAsnGluMetSer 342  
 DB 1447 AAGTACCGCATTTCCGGTTTG-----ACT 1470  
 QY 343 ArgLeuThrGluArgHisLeuThrPhe---LeuAspLeuCysGluGluAsnSerLeuVal 361  
 DB 1471 GCTGTGGCCACTCGGAATTGATTCACAGTAGAT----- 1506  
 QY 362 TyrLysValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeu 381  
 DB 1507 -----GAAAGAAATACTCAGAAA-----TCTGTT 1530  
 QY 382 PheLysIleTyrGluGluAsnLysLysPhe---IleGluPheProHisLeuProLeuVal 400

Db 1531 GTAGAAATCTTCCAGAAACATATGTTTTCGATTCACACACCTCACTACATCAGTCTTG 1590  
Qy LysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGlu 420  
Db 1591 CAAGTTGGGAATCTTAATAGGCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1650  
Qy LysProGlnArgTyrLysAsnArgLysLeuValMetGlnAspLysPheLeuLysArg 440  
Db 1651 ---GGCCAGCGGTATTCCTCAAGAGATTGAATGAGACAGATCACTGCTTTGCTGAAGTT 1707  
Qy 441 AlaThrArgLysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAsp 460  
Db 1708 ACTGTCAGCGCCGATAGATCGAGAAAAGATATCTTACAGACGGTGCACTCAATGAT 1767  
Qy 461 PheSerSerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMet 480  
Db 1768 TATGCTAAAGAT-----AATTATGCTCAAGAGTTTGGCATCAAAATAGTACTTCTCTG 1821  
Qy 481 IleGluCysProGlyLysValLeuLysGluProMetLeu---ValAsnSerValAsnGlu 499  
Db 1822 GCTTCTGTGAGCGCTGATATCTGCTCTCATGCTTAAGTACACAGAGTCTGGAAGG 1881  
Qy 500 GlnIleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValPro 519  
Db 1882 GAAGGGACTTGTCTGCCA-----CAAGTTGGTCAATGAGACATGATGAAT 1926  
Qy 520 GluLysGluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsn---ProCys 538  
Db 1927 AAGAAA-----ATGATCAATGGTGGAAACGCTGAATTAATTGGATCTGC 1968  
Qy 539 Leu-----GluGluAsnAspValValLysPheTyrThrGluLeuIle 552  
Db 1969 ATCAACTTTTCTAGCAAGTGCAGGACAACTAGCGCGTACATTTTGTGAGCAACTGCT 2028  
Qy 553 GlyGlyCysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSer 572  
Db 2029 CAATGTGTAGTATCTGGCATG----- 2052  
Qy 573 IleMetTyrAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThr 592  
Db 2053 -----GCATTAATCCGGAACCACTCTCCACCA 2082  
Qy 593 GlyIleGlyArgPheGlu-----IleAlaAlaThrGluAlaLysAsnMetPhe 608  
Db 2083 GTCAGTGTCTGCGCCTGAGCAAGTAGAAGGTCTTGAAGACTAGATATCATGATGCCACA 2142  
Qy 609 GluArgLeuProAspLysGluGlnLysValLeuMetPheIleIleSerLysArgGln 628  
Db 2143 TCAAAACTCTCCCAAGAAAGAAATGATCTGCTTATGTCATCTGCGCGATATAAT 2202  
Qy 629 LeuAsnAlaTyrGlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGln 648  
Db 2203 CGATCATTTACGGTGATTTGAACCGCATATGTGAGACTGGAACCTGGCATAGTCTCTCAA 2262  
Qy 649 HisIleThrSerGluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySer 668  
Db 2263 TGTGCTGACAAAGCATGCTTTAAGATG-----AGC 2295  
Qy 669 LysArgIlePheTyrGlnIleAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGln 688  
Db 2296 AACAATACATGGCTAATGTTGGCTGAAGATTATGTAAGTTGGAGGAAGAACACA 2355  
Qy 689 GluLeuAspTrpSerGluIleAlaGluIleSerProGluGluLysGluArgLysThr 708  
Db 2356 GTGCTT-----GTTGATGCTCTATCT-----AGCGG 2382  
Qy 709 MetProLeu-----ThrMetTyrValGlyIleAspValThrHisProThr 723  
Db 2383 ATTCTCTTAGTCAGTGATCGACCCACCATATATTGTTGGTGTGATGTTACCCACCTCAC 2442  
Qy 724 SerTyrSerGlyIleAspTyrSerIleAlaAlaValValAlaSerIleAsn---ProGly 742

Db 2443 CCTGAGAGGATTCAAGCCCATCTATTGCTGCTGTTGTGGCATCTCAGGATTGGCCTGAA 2502  
Qy GlyThrIleTyrArgAsnMetIleValThrGln----- 753  
Db 2503 ATCACTAAATATGCTGGATTAGTTTGGCGCTCAAGGCGCATAGCAGAGCTCATTCAGGAT 2562  
Qy 754 -----GluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAsp 771  
Db 2563 CTGTTCAAGAGTGAAGGATCTCAGAAAGGTGGTGACTGGT----- 2607  
Qy 772 IleLeuGluAlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAspAsnArg 791  
Db 2608 -----GGCATCAAAAGGAGTTGCTCATAGCTTCCGTAGATCAACTGGGCATAAA 2658  
Qy 792 AlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgVal 811  
Db 2659 ---CCACTAAGGATCATCTTCTACAGGATGAGTCACTGAGGGACAATTTTACCAGTT 2715  
Qy 812 SerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAsp 831  
Db 2716 TTGCTCTATGAACCTTGATGCCATCGC-----NAGGCCTGCTTCCGCTGGNAGCA 2766  
Qy 832 GlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArg 851  
Db 2767 GGTAT-----CAACCACCACTGACATTTGTGGTGGTGACAGCGCTCATCACAGAG 2820  
Qy 852 LeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGlu 871  
Db 2821 CTGTTTGTCTGACAACCAACATGATCGCATTCGGTGGACAGA----- 2862  
Qy 872 ThrAspValAlaValAlaValLysGlnTyrGluGluAspMetLysGluSerLysGlu 891  
Db 2862 ----- 2862  
Qy 892 ThrGlyIleValAsnProSerSerGlyThrThrValaspLysLeuIleValSerLysTyr 911  
Db 2863 AGTGGGAATATTTTACCT-----GGCACTGTTGTGGACTCTAAATATCGCCACCTACA 2916  
Qy 912 LysPheAspPhePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHis 931  
Db 2917 GAGTTTGATTTTACCTCTGTAGTCATGCTGCTATTCAGGGCAGCTTCTCGACTGCTCAT 2976  
Qy 932 TyrThrValMetTyrAspAspLysGlyMetSerGlnAspLysValTyrLysMetThrTyr 951  
Db 2977 TACCAGTCTCTTGGATGAGAACAACTTACTGCAGATGGACTTCAATCTCGACCAAT 3036  
Qy 952 GlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHis 971  
Db 3037 AACTTATGTTACACGCTATGCAAGATGCACACGCTCAGTTTCAATTGTTCCCTCGCATAT 3096  
Qy 972 TyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyr 991  
Db 3097 TATGCACATCTAGCAGCTTTTAGGGCTCGA----- 3126  
Qy 992 IleGlyAspTyrAlaGlnProArgThr 1000  
Db 3127 -----TTCTACATGAGCCAGACAGA 3147

## RESULT 4

US-11-266-748A-31508  
; Sequence 31508, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and  
; TITLE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; PRIORITY FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03

```

; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 48396
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31508
; LENGTH: 3320
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-31508

Alignment Scores:
Pred. No.: 7,77e-42 Length: 3320
Score: 510.00 Matches: 189
Percent Similarity: 42.8% Conservative: 127
Best Local Similarity: 25.6% Mismatches: 264
Query Match: 9.5% Indels: 160
DB: 26 Gaps: 26

US-10-645-746-3 (1-1020) x US-11-266-748A-31508 (1-3320)

QY 274 PheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuLeuValAspProGln 293
DB 671 TTCTACAAAGCA-----CAACCTGTAATTCAGTTGATGTGTGAAGTTCTTGATATTCAT 724

QY 294 SerCysAsnAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThr 313
DB 725 AATATTGATGGCAACCAAG-----CCT 748

QY 314 IleArgGlnAlaAlaArgProArgIleArgGlnLeuLeuLeuLeuLeuLysLysCys 333
DB 749 CTGACTGATTTCTCGGTAATAATTCACCAAGAGATAAAGGTTTGAAGTTGAAGTG 808

QY 334 AlaGluValTyrAspAsnGluMetSerArgLeuThrGluArgHisLeuThrPheLeu--- 352
DB 809 AC-TCATTGTGGAACAATGAG-----ACGGAA-----ATACCGTGTGTTGTA 849

QY 353 -----AspLeuCysGluGluAenSerLeuValTyrLysValThrGlyLysSerAsp 369
DB 850 TGTAACAGAGCGCTGCGCAGTCATCAACCTTTCTTACAGTTAGAAACCGCCAAAC 909

QY 370 ArgGlyArgAsnAlaLys-LysTyrAspThrThrLeuPheLysIleTyrGluGluAsnLys 389
DB 910 TGTGGAGAGAACACAGTCAGTCAT-----TTACAGAAAAGTATAC 951

QY 389 LysPheIleGluPheProHisLeuValLysValLysSerGlyAlaLysGluTy 409
DB 952 TCTTCAGCTGAAGTACCGCCACCTCTCTGCAAGTCGGCAGCAAGAAACACAC 1011

QY 409 xAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArg11 429
DB 1012 CTACCTGGCCACTAGAGTCTGTAATATT---GTGGCAGGGCAACGATGATATCAAGAGCT 1068

QY 429 eAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLys 449
DB 1069 ACAGACAATCAGACTTCCACTATGATCAAGGCAACAGCAGATCTGCACCATAGACA 1128

QY 449 sGluAenThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheVa 469
DB 1129 AGAGGAAATATGACAGATTGTAAGAGTGCATAATATTAACACAGATCCA-----TTTGT 1182

QY 469 lGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys 489

```

```

Db 1183 TCAGGAGTTTCAATTTAAAGTTCGGGATGAATGGCTCATGTAACTGACGCGTACTTCC 1242
QY 489 sGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrProVal----- 506
Db 1243 AGCACCTATGCTCCAGTATGGAGCAGGAATCGGACAGTAGTACACACCGACCCATGGAGT 1302
QY 507 -----IleArgGlyPheGlnGluLysGlnLeuAsnVal---ValProGluLysGluLe 523
Db 1303 ATGGGACATCGAGG-----AAACAATTCACACAGGAGTTGAAATCAAAATGTG 1353
QY 523 uCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAs 543
Db 1354 GGCTATCGCTTGTGTTTTCACACAG-----AGCAGTGCAGAGAGAA----- 1396
QY 543 pValValLysPheTyrThrGluLeuIle-----GlyGlyCysLysPhe 557
Db 1397 -ATATTGAAGGTTTCACACAGCAGCTGCGTAAGATTTCTAAGATGCGGGATGCCCAT 1455
QY 557 eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAl 577
Db 1456 CCAGGCGCAGCCATGCTTCTGCAAAATATGACAGGGGGCAGACAGCGTA----- 1504
QY 577 aThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPh 597
Db 1504 ----- 1504
QY 597 eGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLy 617
Db 1505 -----GAGCCCATGTTCCGGCATCTCAAGAACACATATTCTGG 1542
QY 617 sValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHi 637
Db 1543 CCTACAGCTTATTATGTCATCTCGCGGGAACACACACAGTAGTGTATCGGAAGTGAACG 1602
QY 637 sTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLy 657
Db 1603 TGTAGGAGACACACTTTTGGGTATGGCTACCAATGTGTTCAAGTCAAGATGTGAATAA 1662
QY 657 sAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLe 677
Db 1663 AACA-----TCTCTCAAACTCTGCAAACTGTGCCT 1695
QY 677 uLysIleAsnAlaLysLeuGlyGlyIleAsnGlnLeuLeuAspTyrSerGluIleAlaGl 697
Db 1696 AAGATAAATGTTAAACTCGGAGGATCAATAATATCTT----- 1735
QY 697 uIleSerProGluLysGluArgLysThrMetProLeuThrMetTyrValGlyI1 717
Db 1736 ----GTACCTCATCAAGACCTTCTGTGTTCCAGCAACCAAGTG---ATCTTTTGGGAGC 1788
QY 717 eAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaValValAl 737
Db 1789 CGATGTCTCATCTCCACCTGCTGCTGATGGAAGAGAGCCCTTCTATGCTGCTGTTGAGG 1848
QY 737 aSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGluCysAr 757
Db 1849 TAGTATGGATGCACACCCCAAGCAGATCTGTGCCACAGTAAGAGTTTCAG----- 1897
QY 757 gProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGlu-----Al 775
Db 1898 -----AGACCCCGACAGGAGATCATCCAGCACTTGGC 1929
QY 775 aLysPheValLys---LeuLeuArgGluPheAlaGluAsnAsnAspAsnAlaProAl 794
Db 1930 CTCCATGTCGGGAACCTTCTATTCAATTTTATAAGTCACTCGGTTCAAG---CCTAC 1986
QY 794 aHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAs 814
Db 1987 TCGTATCATCTTTATCGGGATGCTGTTTCAGAGGGGCGAGTTTATGAGCAGGATATTATTA 2046
QY 814 pGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGlu-A 834

```

```
Db 2047 TGAACCTACTAGCAATTCGAGA-AGCGTCGCATCAGTTT-----GGAGAAAG 2090
Qy 834 sPProGluProLysTyrThrPheIleValIleGluLysArgHisAsnThrArgLeuLeuA 854
Db 2091 ACTATCAACCTGGAATTAACCTACATTTGTAGTTTCAAGAGACATCACACTCGATATTATT- 2149
Qy 854 tGArgMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspV 874
Db 2150 ---TGTGCTGATAGGACAGAAAGGTTGGAAGAGTGGCAATATCCAGCT----- 2197
Qy 874 alalaValAlaAlaValLysGlnTrpGluLysMetLysGluSerLysGluThrGlyI 894
Db 2197 ----- 2197
Qy 894 leValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheA 914
Db 2198 -----GGAACAACAGTTTGATACAGACATTACACACCCATATGAGTTGCG 2240
Qy 914 sPhePheLeuAlaSerHisGleGlyValLeuGlyThrSerArgProGlyHisTyrThrV 934
Db 2241 ATTTTACCTCTGTAGCCATGCTGGAATACAGGGTACCGAGTCGCTTCACACTATCATG 2300
Qy 934 alMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuA 954
Db 2301 TTTTATGGATGATACTGCTTTACTGCGAGATGAACCTTCAGCTGCTTAACCTACCGCT 2360
Qy 954 laPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaH 974
Db 2361 GCCACACTTACGTACGCTGTACACGATCTGTTTCTATACCTGCACCGGCTATTATGCTC 2420
Qy 974 isLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990
Db 2421 ACCTGTGATGATTTAGAGCCAGA---TATCATCTTTGTGGACAAGAACAT 2467
```

## RESULT 5

```
US-11-266-748A-56999
; Sequence 56999, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcriptome Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; PRIOR FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 483996
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 56999
; LENGTH: 3320
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-56999
Alignment Scores:
Pred. No.: 7.77e-42 Length: 3320
```

---

```
Score: 510.00 Matches: 189
Percent Similarity: 42.8% Conservative: 127
Best Local Similarity: 25.6% Mismatches: 264
Query Match: 9.5% Indels: 160
DB: 8 Gaps: 26

US-10-645-746-3 (1-1020) x US-11-266-748A-56999 (1-3320)
Qy 274 PheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuIleValAspProGln 293
Db 671 TTCTACAAGCA-----CAACCTGTAAATTCAGTTTCATGTGTGAAGTCTTGTATATTCAT 724
Qy 294 SerCysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThr 313
Db 725 ATATTGTGAGCAACCAAGA-----CCT 748
Qy 314 IleArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLysCys 333
Db 749 CTGACTGATTCTCATCGGTAATAAATTCACCAAGAGATAAAAGGTTTGAAGGTTGAAGTG 808
Qy 334 AlaGluValTrpAspAsnGluMetSerArgLeuThrGluArgHisLeuThrPheLeu--- 352
Db 809 AC-TCAATTTGTGGAACAATGAG-----ACGGAA-----ATACCGTGTGTTGTTAA 849
Qy 353 -----AspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLysSerAsp 369
Db 850 TGTAAACAAGGAGCGCTGCAGTCATCAAACTTCTCTTACAGTTAGAAAAACGGCCAAAC 909
Qy 370 ArgGlyArgAsnAlaLys-LysTyrAspThrThrLeuPheLysIleTyrGluGluAsnLys 389
Db 910 TGTGAGAGAACAGTAGCGCAGTAT-----TTCAGAGAAAAGATATAC 951
Qy 389 sLysPheIleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTy 409
Db 952 TCTTCAGCTGAAGTACCGCCACCTTCCCTGTCTGCAAGTGGCGGAGGAACAGAAAACAC 1011
Qy 409 rAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArgI 429
Db 1012 CTACCTGCCACTAGAAAGTCTGTAAATATT---GTGCGAGGCAACGATGTATCAAGAAGCT 1068
Qy 429 eAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLy 449
Db 1069 AACAGACAATCAGACTTCCACTATGATCAAGGCAACAGCAAGATCTGCACACAGATAGA 1128
Qy 449 sGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheVa 469
Db 1129 AGAGGAAATTAGCAGATTGGTAAGAAAGTGCAAATTTATGAACAGATCCA-----TTTGT 1182
Qy 469 lGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLy 489
Db 1183 TCAGGAGTTTCAATTTAAAGTTCCGGATGAAATGGCTCATGTAACTGGACGGGTACTTCC 1242
Qy 489 sGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrProVal----- 506
Db 1243 AGCACCTATGCTCCAGTATGGAGGACGGAATCGGACAGTAGCAACACCGGATGAGCT 1302
Qy 507 -----IleArgGlyPheGlnGluLysGlnLeuAsnVal---ValProGluLysGluLe 523
Db 1303 ATGGGACATGCGAGGG-----AAACAATTCACAGAGGATGAAATCAAAATGTGTG 1353
Qy 523 uCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAs 543
Db 1354 GGCTATCGCTGTTGTTTGGCCACACAG-----AGGCAGTGCAGAGAGAA----- 1396
Qy 543 pValValLysPheTyrThrGluLeuLe-----GlyGlyCysLysPhe 557
Db 1397 -ATATTGAAGGGTTTTCACAGACCACTGCGTGAAGATTTCTAAGGATGCGAGGATGCCCAT 1455
Qy 557 eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAl 577
Db 1456 CCAGGCCGCCCATGCTCTCTGCAAAATATGCACAGGGGGGCGACAGCGGTA----- 1504
Qy 577 aThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPh 597
```



Db	922	ATGGACTTCAGATCAAGCTGGCAAGAGCTCAGGGAATCAGATCAAGGTG-----	795	Db	1816	GGGCCCAACGCTG-----	1830
Qy	158	ThrMetIleLeuThrTyrArguyls-----PheHisLeuAenPheSer	172	Qy	510	PheGlnGluLysGlnLeuAenValValProGluLysGluLeuCysCysAlaValPheVal	529
Db	976	-----ACRATTGCGCTGCTGGCCAGCAGATCTGCACCACTGCACGAGTTC	1023	Db	1831	TACCATGACTCT-----GGNAGGAGAAAACTGT-----	1860
Qy	173	ArgGluAenProGluLysAspGluLysAlaAenArgSerTyrLysPheLeuLysAenVal	192	Qy	530	ValAenGluThrAlaGlyAenProCysLeuGluGluAenAspValValLysPheTyrThr	549
Db	1024	ATCGCGCGCGCAGAGGACTCTCAG---CAGCAACATCCAGCACTGGACGTTGTA	1080	Db	1861	-----AATCCAGAGTTGGCAATGGAATATGATCAAT-----AAG	1896
Qy	193	MetThrGlnLysValArgTyrAlaProPheValAenGluLysValGlnPheAla	212	Qy	550	GluLeuIleGlyGlyCys-----LysPheArgGlyIleArg	561
Db	1081	CTG-----AGGAGTCACTCTCTGAATATGTATC-----	1113	Db	1897	AAATGGTGGTGGAGCTGTGTTCAAAGTGCAGCTCGTGAATTTTCACGCATGCAT	1956
Qy	213	LysAenPheValTyrAspAenAenSerIleLeuArgValProGluSerPheHisAspPro	232	Qy	562	IleGlyAlaAenGluAenArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAenGlu	581
Db	1114	-----GTTTCTCGGTGCTTCTACTCCACT	1137	Db	1957	ATTGATGCC---GTGCACAGACTATCGCGCAACTAGTTTATACATGCAATGTAATGGC	2013
Qy	233	AsnArgPheGlnGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyr	252	Qy	582	TyrAlaPheTyrLysAenCysThrLeuAenThrGlyIleGlyArgPheGluIleAlaAla	601
Db	1138	ATGTTGGCCGCAAA---GACATTGGTGTATGGCTGGAATGCTGGAAAGGATACTAT	1191	Db	2014	ATGGTTTTCAATGAATGCCAGATAGAGTGGGGTCAAGTGTCTCTTAATTAACATCGAA	2073
Qy	253	IleGlyIleLysGluLeuPheAspGlyGluProValLeuAenPheAlaIleValaAspLys	272	Qy	602	ThrGluAlaLysAenMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPhe	621
Db	1192	CAGAGCTCGGCCCACTCAGATGGACCTCA---TTGAATAGACATATCCTCTACC	1248	Db	2074	GCTGCCCTGAGCAACATTTCACACAAGGGCTCTCTCAA-----CTCCAGCTG	2118
Qy	273	LeuPheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeu---LeuIleValaAsp	291	Qy	622	IleIleIleSerLysArgGlnLeuAenAla---TyrGlyPheValLysHisTyrCysAsp	640
Db	1249	CCATTCTTCAAA---CCTATCAGCGTGGTAGAATATGTCAAGAAATTGCTGGGCACA	1302	Db	2119	CTCATTTGATTTCTCCAGATGTTAATGGTATTATGAAGAATTAAGAGGGTGTGTGAG	2178
Qy	292	ProGlnSerCysAen---AspAspValArgLysAspLeuLysThrLysLeuMetAlaGly	310	Qy	641	HisThrIleGlyValAlaAenGlnHisIleThrSerGluThrValThrLysAlaLeuAla	660
Db	1303	CTTACTAATGCTAATGGCCCTGACCTAGCGGCTCTTTCT-----	1344	Db	2179	ACTGAACCTTGGATAGTATCCAGTGCCTCAAGCCA-----GGCGCAAGCTCTTG	2229
Qy	311	LysMetThrIleArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAenLys	330	Qy	661	SerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAen	680
Db	1345	-----GACATTGATCGCTGAGAGTTAAGAAAGCACTACGGGAGTTGCT	1389	Db	2230	AGCTTA-----GACAGGCAGTTCTCTGGAAAAATGTCTCACTCAAAATCAAT	2274
Qy	331	LeuLysCysAlaGluValTrpAspAenGluMetSerArgLeuThrGluArgHisLeuThr	350	Qy	681	AlaLysLeuGlyGlyIleAenGlnGluLeuAspTrpSerGluIleAlaGluLysPro	700
Db	1390	GTTGAACAACACACACAGGGGAAGCAGCAAGTAGACAGTCACTACG---ATTACA	1443	Db	2275	GTCAAGGCTGGAGGACGCAACTCAGTTCTT-----LeuThrMetTyrVal	715
Qy	351	PheLeuAspLeuCysGluGlnAenSerLeuValTyrLysValThrGlyLysSerAspArg	370	Db	2305	-----CAGAGACCTCTGTACCGGTGGGTGGAAACACACAATAATTTT	2352
Db	1444	TCAGGCCACTGAGTCAGCTGAAC-----TTTCTATGATGGAACTACCCAG---	1491	Qy	716	GlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaVal	735
Qy	371	GlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGluGluAenLysLys	390	Db	2353	GGTCCGATGTCAACCATCTCTCTCGAGAGGACTCATCGGCTCGATCGCAGCTGTG	2412
Db	1492	-----ACTGTTATTCAGTACTTCTCGCAGCGGTACAAA	1524	Qy	736	ValAlaSerIleAen---ProGlyGlyThrIleTyrArgAenMetIle-----	750
Qy	391	Phe---IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTyr	409	Db	2413	GTGGCTCCATGGAGCTGCTGAGATCACAAGTACAAGCCCTCTCTCTGCCAGCCA	2472
Db	1525	TACAGCTGCAGTACAGCTCTTGGCCCTCTCTGCAATCCGCAACCTTCTTAACCTATA	1584	Qy	751	-----ValThrGlnGluCys	756
Qy	410	AlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAenArgIle	429	Db	2473	CCTGGCAGGAGATTATACAAGATCTCTTACCATGACTGAAGTTCGCGAGAATGCTGAT	2532
Db	1585	TATTTGCCAATGGAGTATGCACCAATTGTAGAA---GGCAAGAGATACTCCAAGAAGCTC	1641	Qy	757	ArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluLys	776
Qy	430	AspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLys	449	Db	2533	GCTCCAGCACAGAG---GCTGAAGTTTCAAGAAGAAATTCATATCGCGCGAATG	2586
Db	1642	ATNGACAAACAGTACTGGCTCTCTGAGAGCAACATGCCACCTCCCCAGAAAGGGAG	1701	Qy	777	PheValLysLeuLeuArgGluPheAlaGluAenAenAspAenArgAlaProAlaHisIle	796
Qy	450	GluAenThrLeuLysMetLysLysGluLeuAspPheSerSerGluGluLeuAenPheVal	469	Db	2587	TTCAAGGAGTGTCTTATCTTCTTACAGTAAGAAATGCTAAACGTAAGCCTCAAGGATA	2646
Db	1702	CAGAAAAATCATTGAGATGTTTCAACACACAACACTACCCCGCTGATAAGGTG-----GTG	1755	Qy	797	ValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeu	816
Qy	470	GluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys	489	Db	2647	ATATTTTACAGGATGGTGAAGTATGAGTATGACAAATTCCTTCTGCTCTATGAGATG	2706
Db	1756	ACTGATTTTGAATTAATATTTCCAATCAGATGGGCCACTATGCCAGCTGCGCTGCTGCT	1815	Qy	817	ArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPro---	835
Qy	490	GluProMetLeuValAenSerValAenGluGlnIleLysMetThrProValIleArgGly	509	Db	2707	GACGCAATCAAGAGGCTATTGTCATCTTTG-----GACCCAGCA	2745







Qy	416	uGluValHisGluLysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLy	436
Db	1517	CAAAATAGTGGA--GGACAGAGATACTCCAAGAGATTAAATCAGAATCAGATAAGAGC	1573
Qy	436	sPheLeuLysArgAlaThrArgLysProHisAspTyrLysGlnuThrLeuLysMetLe	456
Db	1574	TCTTTTGGAGAGACATGCCAACACCCAGTGGAGCGGTATATTAAGATGGT	1633
Qy	456	uLysGluLeuAspPheSerSerGluGluLeuAsnPheValGluArgPheGlyLeuCys	476
Db	1634	TAAACATAACGCTTATCAGGATGATCTCT--TACGCAAAAGATTGGCATTAAAGT	1687
Qy	476	rLysLeuGlnMetIleGluCysProGlyLysValLeuLysGluProMetMet---ValAs	495
Db	1688	AAGTGATCGTGTGCATCAGTAGAGGACAGAAATTTTACCGCTCCACGGCTTAAGTACAA	1747
Qy	495	nSerValAsnGluGlnIleLysMetThrProValIle-----ArgGlyPheGlnG	512
Db	1748	TGAGACTGGTCGAGAGAGGATTGCTTACCAAGAGTTGGTCATGGAAATATGATGAACAA	1807
Qy	512	uLysGlnLeuAsnValValProGluLysGluLeuCysCysAlaValPheValValAsnG	532
Db	1808	GAATAATGGTAAATGGTGTAAAGTCAGAAAGTCGATGTGCTCAATTTTGCCCGTAAT--	1865
Qy	532	uThrAlaGlyAsnProCysLeuGluGluLeuAsnAspValValLysPheTyrThrGluLeu	552
Db	1866	-----GTGAGGAAAGTTGTTGCTGGGTCTGCGCATGAACCTGTC	1906
Qy	552	eGlyGlyCysLysPheArgGlyIle-----ArgIleG	563
Db	1907	TCTGATGTCTCAAGCCTCAGGAATGGATTTCGCTCCAGAACCTATCTTCCACCACATAA	1966
Qy	563	yAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAsnGluTyrAl	583
Db	1967	TGCATCTCCAGATCAA-----GTGAGCGGTGCTCTAAAGCTAGGTAT--	2009
Qy	583	aPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluLeuAlaAlaThrG	603
Db	2010	-----CATGA	2014
Qy	603	uAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPheIle	623
Db	2015	TGCAATGAACCTTCTTGA-----CCACAGCGCAGGAACTTGATTGCTATTGGGAT	2068
Qy	623	eIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAspHisThrI	643
Db	2069	ACTACCTGATAACATGGCTCGCTTTATGGTGATTGAAGCGCGTGTGAATATGATCT	2128
Qy	643	eGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeuAr	663
Db	2129	TGGAATAGTTTCACAACTGCTTTCACAGACAGCAGGTGTTCAAAATG-----	2174
Qy	663	gHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsnAlaLysLe	683
Db	2175	-----AACAAACAAATCTTGCAAAATCTTGCTCTGAAGATAAATGTCAAGGT	2221
Qy	683	uGlyGlyIleAsnGlnGluLeuAspTyrSerGluIleAlaGluIleSerProGluGluLy	703
Db	2222	TGGGGGCGGAGAACACTGTGCTG-----GTGGATGCTGTGCA-----	2258
Qy	703	gGluArgLysThrMetProLeu-----ThrMetTyrValGlyIleAs	718
Db	2259	-----AGGGGTATTCTCTGGTAAACCGACAGACCTACAAATATATTTGGTGTGTA	2308
Qy	718	pValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValValAlaSe	738
Db	2309	TGTTATCCCATCCTCATCTCGAGAGGACAGCAGTCCCTCAATCTCTGTGTGAGCCTC	2368
Qy	738	rIleAsn---ProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluCysAr	757
Db	2369	CCAAGATTGGCTCAGGTGACAAAGATATGCTGGGTGCTGTTCTGCTCAA-----	2417

Qy	757	gProGlycIuAArgAlaValAlaHisGlyAArgGlu-----	768
Db	2418	-----GCCACCACCAAGAGCTGATAGAAGATCTATATAAAATCTG	2458
Qy	769	-----ArgThrAspIleLeuGluAlaLysPheValIysLeuArgGluPh	784
Db	2459	GCAGGATCCAGAGAGAACAGTTAGTGGTGGCATGATCCGTGAGCTGCTTATATCTT	2518
Qy	784	eAlaGluAsnAsnAspAsnArgAlaProAlaHisIleValIysTyrArgAspGlyValse	804
Db	2519	CAAAAGATCAACTGGTGAGAAG--CCCCAGCGAATAATATTTTACAGGCGATGGCGTTAG	2575
Qy	804	rAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLy	824
Db	2576	TGAAGGCCAATTTTACCAAGTTCTACTTTTATGAATTTGAAATGCAATCCGAAAAAGCATGTGC	2635
Qy	824	sGlnPheMetSerGluArgAspGlyGluAspProGluProLysTyrThrPheIleValIi	844
Db	2636	CTCCCTGGAGACA-----AATTACCACCAAGGTGACTTTCATTGTGGT	2680
Qy	844	eGlnLysArgHisAsnThrArgLeuLeuArgArgMetGluLysAspLysProValValas	864
Db	2681	TCAGAAACGTCACCAACAAGATATTATTGGACATAATCACACGATCAGAACTCAGTTGA	2740
Qy	864	nLysAspLeuThrProAlaGluThrAspValAlaValAlaLysGlnTrpGluGl	884
Db	2741	CAGG-----	2744
Qy	884	uAspMetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValas	904
Db	2745	-----AGCGGGACATCACTCCCT-----GGTACGGTTGTAGA	2776
Qy	904	pLysLeuIleValSerLysTyrLysPheAspPheLeuAlaSerHisGlyValle	924
Db	2777	TTCAAAGATCGTCATCAACTCAGTTTGACTTCTACCTGTGTAGCCATCTCTGGCATTA	2836
Qy	924	uGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnas	944
Db	2837	GGGTACTAGTCGTCACCTCATTCATGTCCTTGTGGGATGAACAACTTCACAGCTGA	2896
Qy	944	pGluValItyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIi	964
Db	2897	TGCATTGCAGATTCTTACCAACAACCTTTGCTACACCTATGCAAGGTGCACTCGCTCTGT	2956
Qy	964	eSerLeuProValIProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuty	984
Db	2957	ATCAAAATGTTCCACCTCGTTATTATGCTCATCTGGCTGCCTCCGTGCT-----	3005
Qy	984	rArgThrTyrLysGluHisTyrIleGlyAspTyrAlaGln-----ProArgth	1000
Db	3006	-CGTTTCTATATGNAACCAACATACATCTGACAGCAGCTCTGTCTAGTGGGCTGGTGT	3064
Qy	1000	rArgHisGluMetGluHisPheLeuGlnThrAsnValIysTyrProGly	1016
Db	3065	ACGTGGGGCCACTTCTGGCTCATCAACATCAACGCTACTCGGGCCCCCTGGT	3113

## RESULT 8

RES001 8  
US-11-218-305-22486  
; Sequence 22486, Application US/11218305  
; Publication No. US20060141495A1  
; GENERAL INFORMATION: MONSANTO TECHNOLOGY, LLC  
; APPLICANT: MCLAIRD, Paul L.  
; APPLICANT: Tao, Kunsheng  
; APPLICANT: Wu, Nensheng  
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
; TITLE OF INVENTION: Corn.  
; FILE REFERENCE: 38-21 (53660)B  
; CURRENT APPLICATION NUMBER: US/11/218,305  
; CURRENT FILING DATE: 2005-09-01  
; PRIOR APPLICATION NUMBER: US 60/606,880  
; PRIOR FILING DATE: 2004-09-01  
; NUMBER OF SEQ ID NOS: 25043



Db	2518	-----AAACAAGCAGCAGTATCTTGC	AAATGTCCTCGAATAAATATGTAAG	2565
Qy	683	LeuGlyGlyIleAanGlnGluLeuA	spTrpSerGluIleAlaGluIleSerProGluGlu	702
Db	2566	GTTCGGGGAGGAATACTGTCCT	-----GTTGATGCT	2598
Qy	703	LysGluArgArgLysThrMetProLeu	-----ThrMetTyrValGlyIle	717
Db	2599	TTGCAAGGAGA-----ATCC	CTTGTAGTCACATAGCAGACTATATCTTTGGTGCT	2652
Qy	718	AspValThrHisProThrSerTyr	SerGlyIleAspTyrSerIleAlaAlaValValAla	737
Db	2653	GATGTGACCCATCCCATCTCGG	GGAAGATTCTAGTCTTCCATTGCAGCTGTGGTTGCT	2712
Qy	738	SerIleAan---ProGlyGlyThr	IleTyrArgAanMetIleValThrGln-----	753
Db	2713	TCTCAAGACTGGCCTGAGGTTT	ACAAAGTATGCAGGATTGTGTAGTGTCAAGCCCATCGT	2772
Qy	754	GluGluCysArgProGlyGluA	ArgAlaValAlaHisGlyArgGluAArgThrAspIleLeu	773
Db	2773	CAAGATTGATACAGGATCTTTT	CAAGGTATGGCAGAGTCCGGAAGGGGAGCTGTCTCT	2832
Qy	774	GluAlaLysPheValLysLeu	ArgGluPheAlaGluAanAanAspAanArgAlaPro	793
Db	2833	GGTGGCATGATCAGGAGCTTCT	ATATCTTCTGGAGGGCAACTGCACAGAAA---CCA	2889
Qy	794	AlaHisIleValValTyrArg	AspGlyValSerAspSerGluMetLeuArgValSerHis	813
Db	2890	AAGAGGATCATATTCTACAGG	ATGCGGTGCTAGTCAGGGACAAATCTACCAAGTTCTGTTG	2949
Qy	814	AspGluLeuArgSerLeuVal	SerGluValLysGlnPheMetSerGluAArgAspGlyGlu	833
Db	2950	TATGAACCTGATGCCATTAGA	AGGCCTCTGCGTCATTTGGAGTCT-----	2994
Qy	834	AspProGluProLysTyrThr	PheIleValIleGlnLysArgHisAenThrArgLeuLeu	853
Db	2995	GACTACCAAGCCTCCAGTTACT	TTTGTGCTGTCTCCAGAGGCTCATCACACGAGTTGTTT	3054
Qy	854	ArgArgMetGluLysAspLys	ProValValAanLysAspLeuThrProAlaGluThrAsp	873
Db	3055	GCTAATATCAATGATTAATCG	TGCTGATAAA-----	3090
Qy	874	ValAlaValAlaAlaValLys	GlnTrpGluAspMetLysGluSerLysGluThrGly	893
Db	3091	-----	AGCGGG	3096
Qy	894	IleValAanProSerSerGly	ThrThrValAspLysLysIleValSerLysTyrLysPhe	913
Db	3097	AACATATCGCT-----GG	CACCGTGTGGACTCGAAGATCTGCCATCAACATGAGTTT	3150
Qy	914	AspPhePheLeuAlaSerHis	HisGlyValLeuGlyThrSerArgProGlyHisTyrThr	933
Db	3151	GATTTCTACCTGTGCAGCCAT	GTCTGGCATTCAGGGAACAAGCCGCTGTGCCATTACCAT	3210
Qy	934	ValMetTyrAspAspLysGly	MetSerGlnAspGluValTyrLysMetThrTyrGlyLeu	953
Db	3211	GTTCTGTGGGATGAGAACAA	CTTTACCGCTGATGGGTTGCAAACTCTCACCAACAACATTG	3270
Qy	954	AlaPheLeuSerAlaArgCys	ArgLysProIleSerLeuProValProValHisTyrAla	973
Db	3271	TGTTACAGATGTTAGTGCA	CAGCTTCAGTATCGATTGTCTCTCTGCTACTATGCT	3330
Qy	974	HisLeuSerCysGluLysAla	LysGluLeuTyrArgThrTyrLysGluHisTyrIleGly	993
Db	3331	CACCTGGCAGCCTTCCGAGCT	-----CGGTTCTACATGGAGCCAGATACGAGT	3378
Qy	994	AspTyrAlaGlnProArgThr	Arg	1001
Db	3379	GACAGTGATCTATGGCAAG	CCGT	3402

RESULT 9  
US-11-266-748A-25558

```

; Sequence 25558, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcriptome Microarray Technology and
; FILE REFERENCES: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; CURRENT FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 483996
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25558
; LENGTH: 1793
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-25558

Alignment Scores:
Pred. No.: 1.73e-40 Length: 1793
Score: 493.00 Matches: 160
Percent Similarity: 43.7% Conservative: 89
Best Local Similarity: 28.1% Mismatches: 202
Query Match: 9.2% Indels: 120
DB: 8 Gaps: 18

US-10-645-746-3 (1-1020) x US-11-266-748A-25558 (1-1793)

Qy 434 GlnAspLysPheLeuLysArgAlaThr---ArgLysProHisAspYr
||| :||: |||||
Db 3 CAGACCTCAACCATGATCAGAGCAACTGCTAGTTCGCGCCCGCATCGG

Qy 453 LeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPhe
|||:|||||:||||:||||:||||:||||:||||:
Db 63 AGCAAAATTGATCGGAAGTGCAGAGTTTCAACACAGATCCA-----TAC

Qy 473 GlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeu
|||:||||:||||:||||:||||:||||:||||:
Db 117 GGAATCATGCTCAAGATGAGATGACAGCATGACTGGCGGGTGCTG

Qy 493 LeuValAsnSerValAsnGluGlnIleLysMetThrProVal-----
:||||: :||:
Db 177 ATCTCTACGGGGCAGGAATAAAGCTATTTCGCACCCCTGTCACGGGC

Qy 511 GlnGluLysGlnLeuAsnVal-----ValProGluLysGluLeuCysCys
:||||:||||:||||:||||:
Db 237 CGGAACAAGCAGTTCCACACCGGCATCGAGATCAAGGTGTGGGCCATT

Qy 530 ValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAspValVal
:||||:||||:||||:||||:
Db 297 CCCAG-----CCCCAGTGCACGGAAGTCCAT-----CTG

Qy 550 GluLeuIle-----GlyGly-CysLysPheArgGln
:||||:
Db 339 GACGAGCTCAGAAGATCTCGAGAGACGCTGGCATGCCATCCATCAGCGGC

```

```
QY 563 YAlaAsnGluAsnArgGlyAlaGlnSerIle-----MetTyrAspAlaThrLysAsnGlu 581
Db 399 TGAATAATACCGC-CAGGGGGCGGACGCGTGGAGCCCATGTTCGGCACCTGGAAGAACAC 457
QY 581 uTyrAlaPheTyrLysAsnCyThrLeuAsnThrGlyIleGlyArgPheGluIleAlaAl 601
Db 458 GTATGCG----- 464
QY 601 aThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLysValLeuMetPh 621
Db 465 -----GGCCTGCAGCTGGT 478
QY 621 eIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAspHi 641
Db 479 GGTGTCATCTCTCCCGCAGACGCGCGGTGTAGCCGAGGTCAAGCGGTGGAGACAC 538
QY 641 sThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAlaSe 661
Db 539 GGTGCTGGGATGGCCAGCGAGTGGCGTGAGATCAAGAACGTG----- 581
QY 661 rLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsnAl 681
Db 582 -----CAGAGGACCAACGCGCACAGACCTGTCCAACCTCTGCGCTGAAGATCAACGT 631
QY 681 aLysLeuGlyGlyIleAsnGlnGluLeuAspTyrSerGluIleAlaGluIleSerProGl 701
Db 632 CAAGCTGGGAGGCGTGAACAACATCTCTG-----CTGCCCCA 667
QY 701 uGluLysGluArgLysThrMetProLeuThrMetTyrValGlyIleAspValThrHi 721
Db 668 GGCAGCGCCCGGTGTTCAGCAGCCGCTC-----ATCTTCTGGAGCAGACGTCACTCA 724
QY 721 sProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValAlaSerIleAsnPr 741
Db 725 CCCCCCGCGCGGATGGGAAGAAGCCCTCCATTGCCGCGTGTGGCGCAGCATGAGCGC 784
QY 741 oGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluGluCysArgProGlyGluAr 761
Db 785 CCACCCCAATCGCTACTGCGCCACCGCTGCGCGTCAGCAG----- 824
QY 761 gAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLe 781
Db 825 -----CACCGCAGAGATCATACAGACCTTGGCGCCCATATGTCGCGAGCTCCT 874
QY 781 uArgGluPheAlaGluAsnAsnAspAsnArgAlaProAlaHisIleValValTyrArgAs 801
Db 875 CATCCAGTCTACAGTCCACGCGCTCAAG---CCACCCGCGATCATCTTCTACCGCGA 931
QY 801 pGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLysSe 821
Db 932 CGGTGTCTCTGAAGGCGAGTTCACGAGGTCTCCACACGAGTGTGTCGCCATCCGTGA 991
QY 821 rGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluPyrLysTyrThrPh 841
Db 992 GGCCTGTATCACTA-----GAAAGAAGACTACACCGCCGGGATCACTT 1036
QY 841 eIleValIleGlnLysArgHisAsnThrArgLeuLeuArgArgMetGluLysAspLysPr 861
Db 1037 CATCGTGGTGACAGAGGACCAACCCCGGCTCTTC---TGCACTGCACAAGAACGAGCG 1093
QY 861 oValValAsnLysAspLeuThrProAlaGluThrAspValAlaValAlaValLysGlu 881
Db 1094 GGTGGGAAAGTGGAAACATTCACGCA----- 1121
QY 881 nTrpGluGluAspMetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyTh 901
Db 1122 -----GGCAC 1126
QY 901 rThrValAspLysLeuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHisHi 921
Db 1127 GACTGTGGACACGAAATATCAACCCACCCCGAGTTCGACTTCTACCTGTGTAGTACCGC 1186
QY 921 sGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMe 941
```

```
Db 1187 TGGCATCCAGGGGACAGCAGCGCTTCGCACTATACGTCCTCTGGGACGACAATCGTTT 1246
QY 941 tSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaAArgCysAr 961
Db 1247 CTCTCTGATGAGCTGCAGATCCCTAACCTACCACTGTGTGCACACCTAGTGCCTGCAC 1306
QY 961 gLysProIleSerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLy 981
Db 1307 ACGTCTCGTCTCCATCCAGCGCAGCATACTAGGCTCACCTGTGTGGCTTCGCGGCCAG 1366
QY 981 sGluLeuTyrArgThrTyrLysGluHis 990
Db 1367 G---TACCACCTGGTGGATAAGAACAT 1391

RESULT 10
US-10-449-902-19320
; Sequence 19320, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19320
; LENGTH: 2503
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK069685
; DATABASE ENTRY DATE: 2001-12-06
US-10-449-902-19320

Alignment Scores:
Pred. No.: 1,46e-39 Length: 2503
Score: 486.00 Matches: 199
Percent Similarity: 40.6% Conservative: 116
Best Local Similarity: 25.7% Mismatches: 289
Query Match: 9.1% Indels: 171
DB: 32 Gaps: 32

US-10-645-746-3 (1-1020) x US-10-449-902-19320 (1-2503)
QY 238 SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyIleLysGlu 257
Db 220 TCGAGGGAATCTGCACAAAGGCGCATCTTTGCGTGAAGGGGTTCTACAGAGTTCCCGGTC 279
QY 258 LeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAla 277
Db 280 ACGCAGCAGCGCTTCTCT---CTGAACATAGACATGTCTTCGACTGCTTTC----- 327
QY 278 ProLysMetSerLeuLeuAspTyrLeuLeuIleValAspProGlnSerCysAsnAsp 297
Db 328 -----ATTGACCTGTCGGGTGCTGAAT 351
QY 298 AspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAla 317
Db 352 TTGTGAGAAAAGCTATTTCGACGTCGAATC---ACTAATGCTATTACTGTGGGATATTTT 408
QY 318 AlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLysCysAlaGluValTrp 337
Db 409 TTGAACAATTTATGGGAATGAATTGATGAGACCCCTTAAGGCTGTTAAGTTGAATC--- 465
```

Qy	338	AspAenGluMetSerArgLeuThrGluArgHisLeuThrPheLeuAapLeuCysGluGlu	357
Db	466	---ACTCACCAGGAATACTACCGAAGAAGTAC---AGCATTTGGCTGCCTCCACCGCAGCAG	519
Qy	358	AasnSerLeuValTyrlYsvalThrglySserAapArgglyArGAsnAlalalyeLysTyx	377
Db	520	TCTCGAGATGTTCAGAGCTTCACA---TCATCTCATGGTATCAAGAGCTGTCAAGAGGATAT	576
Qy	378	AspThrThrLeuPheLysIleTyrlGluGluAasnLysLysPhe-----IleGluPhe	394
Db	577	-----TTCAACAAAAANTACATCTGAAGTTAGCTTTT	609
Qy	395	ProHisLeuProLeuValLysValYsSetGlyAlalYsGluTyrlAlaValPrometGlu	414
Db	610	GATTATCTTCATGCCCTGCAAGTTGGCAGCAAGGAGAGACCGAAATTAACCTGCCCATGGAG	669
Qy	415	HisLeuGluValHisGluLysProGlnArgTyrlYsAasnArgIleAapLeuValmetGln	434
Db	670	CTTTGGCAATATA---GTTTCCTGGACAACGATACAAAGAACAGGCTCAGTCCGACACAGGCT	726
Qy	435	AspLysPheLeuLysArgAlaThrArgLysProHisAaspTyrlYsGluAasnThrLeuLys	454
Db	727	TCCAATCTGATTAAACATAACCAACGATCGTCTTGTCAC---CGTAGAGCTCCATTCGT	783
Qy	455	MetLeuLysGluLeuAapPheSerSerGluGluLeuAasnPheValGluArg-----	471
Db	784	CAG-----ACTGTTAGCAGCAACCAGTATAACAGTAGTACGGAACGCGCAGATGAG	831
Qy	472	PheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLysGluPro	491
Db	832	TTTTGGCATAGAAAGTTGACTCTTATCTCTACTTCTTAAAGGCTACAGATTTTGAAGCTCCA	891
Qy	492	MetLeuValAasnSerValAasnGluGlnIleLysMet---ThrProValIleArgGlyPhe	510
Db	892	ATGCTGAAGTACCATGATTTCTGGAAGGGTGAGAGATATCGACGCCACGAGGATGGGGCGTGG	951
Qy	511	GlnGluLysGlnLeuAasnValVal-----ProGluLysGluLeuCysCysAlaVal	527
Db	952	AACATGAAAGACAAGAAAGTAGTTAAACGGTCTACAATTAAGCTGGGCATGT-----	1005
Qy	528	PheValVallAasnGluThrAlaGlyAasnProCysLeuGluGluAasnAapValValLysPhe	547
Db	1006	-----GTCACACTTGTGCGAGGGT-----TTGGATAATCGTGTGTGAAGCAATTC	1050
Qy	548	TyrThrGluLeuIleGlyGlyCySLysPheArgGlyIleArgIleGlyAlaAasnGluAasn	567
Db	1051	TGCCTTCATTTGGTCAGAACGTCCAAAATAACTCGACTGGAC-----	1092
Qy	568	ArgGlyAlaGlnSerIleMetTyrAspAlaThrLysAasnGluTyrlAlaPheTyrlLysAasn	587
Db	1093	-----TTTGCGAAT	1101
Qy	588	CysThrLeuAasnThrGlyIleGlyArgPheGluIleAlalaThrGluAlaLysAasnMet	607
Db	1102	GTGAGCCTT-----CCAAATATTGAAGCTGATCTCTCAATAATGTT	1140
Qy	608	PheGluArgLeuPro-----AapLysGluGlnLysVal	618
Db	1141	AAAACTGATCTCTTATGCGGTATCAGGAAGCATGCAGCTGGTCGAGGGATAAACAAGATT	1200
Qy	619	LeuMetPheIleIleIle-----SerLysArgGlnLeuAasnAlaTyrlGlyPheVal	635
Db	1201	GACCTCTACTTGTGTAAATCACAGATGATANAATAATGCCACGTTATATGGTGACGTT	1260
Qy	636	LysHisTyrlCysAapHisThrIleGlyValAlaAasnGlnHisIleThrSerGluThrVal	655
Db	1261	AAAAAGAATCTGTGAACACAGAAATCGGTGTATTGTTCACAGTGTGTCTCAGCGCAAGCAAGTC	1320
Qy	656	ThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyx-----	673
Db	1321	TACAAG-----GAGAGGAATGTTTCAGTATCTGCGCA	1350
Qy	674	GlnIleAlaLeuLysIleAasnAlalyeLysLeuGlyGly-----IleAasnGln	688

RESIT.T 11

US-11-218-305-543  
: Sequence 543. Application US/11218305

; sequence 543, Application US/1  
 ; Publication No. US20060141495A1  
 : GENERAL INFORMATION:

Db	1351	AAATGTTGCTCTTAAAGATCAATGCGCAAGGCTGGAGGAAGAACTCGGTATTTCCTTAATGTA	1410
Qy	689	GlulLeuAspTrpSerGluIleAlaGluIleSerProGluGluLysGluArgAglYsThr	708
Db	1411	GAGCAAGTTTACCGGGTGTTCAAAG--AGCCCA-----	1443
Qy	709	MetProLeuThrMetTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIle	728
Db	1444	-----ACTATTATATTGGTGTGTGTTTACCATCTCTGGTCTTGTGATAAAGT	1494
Qy	729	AspTyrSerIleAlaValAlaSerIleAsn---ProGlyGlyThrIleTyrArg	747
Db	1495	ACCCCTCCAAATGCTTCGGTGTGTCTCGCAGACTGGCTGAGTGACCAAGTATAAT	1554
Qy	748	AsnMetIleValThrGlnGluGluCysArgProGlyGluArgAlaValAlaHisGlyArg	767
Db	1555	TCTGTGTT-----CGTATCAAGCTTCTCGTAAG	1584
Qy	768	GluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAlaGluAsn	787
Db	1585	GAGATTATACAAGATCTTGATAGACTGTTAGGGAACCTTCTCAATGCATTTC-----AAAAGG	1641
Qy	788	AsnAspAsnArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGlu	807
Db	1642	GATCCCAAGATGGAGCGGAAGCAGCTCATTTCTCAGGACCGCTTAAGCGAGGCTCAG	1701
Qy	808	MetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMet	827
Db	1702	TTCAGCAAGTTGTAGAGCGAANAATACCGAGATAGAAAAAGGCTTGGAAAGTCTCTGTAT	1761
Qy	828	SerGluArgAspGlyGluAspProGluPurProLysTyrThrPheIleValIleGlnLysArg	847
Db	1762	CGTGGCAAG-----CCACGAATTACCTTTCATAGTGGTGCAGAAAGAGG	1803
Qy	848	HisAsnThrArgLeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeu	867
Db	1804	CATCATACAAGGCTGTTCACCAACATTACATGAT-----	1839
Qy	868	ThrProAlaGluThrAspValAlaValAlaLysGlnTrpGluGluAspMetLys	887
Db	1840	---CCACGCGCATGGAT-----	1854
Qy	888	GluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeuIle	907
Db	1855	-----GGGACTGGAATGTTCTGCA-----GGCACAGTAGTTGATCAGTGATC	1899
Qy	908	ValSerLysTyrLysPheAspPheLeu-AlaSerHisHisGlyValLeuGlyThrSer	927
Db	1900	TGTCACCCCTCGAGAGTTTGATTTCTTCTGTGCAGCCAAAGCCGGGATCAAGGGACAAG	1959
Qy	927	ArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTy	947
Db	1960	CCGTCTCTAGCAATTCACATGTGCTGGCGCAGCAACAACCTTCACGCGCATCAGGTTCA	2019
Qy	947	xLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuPr	967
Db	2020	GTCTGTCACAAACAACTGTGCTACTTATATACAAAGCTGCACCTCGCTCGGTGCTATTCC	2079
Qy	967	oValProValHisTyrAlaHisLeuSerCysGluLysAlaLys	981
Db	2080	ACCTCCGTGTTTACTACGCTCATAGCTCGCATTCGCGCTCGT	2122

RESULT 11  
US-11-218-305-543  
; Sequence 543, Application US/11218305  
; Publication No. US2006014195A1  
; GENERAL INFORMATION:  
; APPLICANT: MONSANTO TECHNOLOGY, LLC  
; APPLICANT: McLaIRD, Paul L.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Wu, Kunsheng  
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping

```

, TITLE OF INVENTION: Corn.
, FILE REFERENCE: 38-21 (53660)B
, CURRENT APPLICATION NUMBER: US/11/218,305
, CURRENT FILING DATE: 2005-09-01
, PRIOR APPLICATION NUMBER: US 60/606,880
, PRIOR FILING DATE: 2004-09-01
, NUMBER OF SEQ ID NOS: 25043
, SOFTWARE: PatentIn version 3.2
, SEQ ID NO 543
, LENGTH: 5408
, TYPE: DNA
, ORGANISM: Zea mays
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (766)..(767)
, OTHER INFORMATION: n is a, c, g, or t
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (769)..(769)
, OTHER INFORMATION: n is a, c, g, or t
US-11-218-305-543

```

Alignment Scores:		
Pred. No.:	2.32e-38	Length:
Score:	479.00	Matches:
Percent Similarity:	44.3%	Conservative:
Best Local Similarity:	27.3%	Mismatches:
Query Match:	9.0%	Indels:
DB:	7	Gaps:
		22
		5408

US-10-645-746-3 (1-1020) x US-11-218-305-543 (1-5408)

Qy	470	GluArgPheGlyLeu	CysSer	Leu	GlnMet	IleGlu	CysPro	GlyLeu	ValLeu	Leu	489
Db	2082	CAGCAATTTGGGATTTCTCTTGGATGACGAATGACAGACAGACCGGTAGGATCTCTCTCT									2141
Qy	490	GluProMetLeu	---Val	AsnSer	ValAsn	GluGln	IleLeu	LeuMet	ThrPro	ValIle	Arg 508
Db	2142	CCACCAAACTCAAACTTGGGGCATCCAAATGGGCAGACCTCCAAATTGATGATATCGATCAT									2201
Qy	509	GlyPheGln	-----Glu	LeuSer	GlnLeu	AsnVal	ValPro	GluLeu	ValSer	GluLeu	Cys 524
Db	2202	GGCTGCAGTGAATCTTTGTGAAGAGAGACTA-----GTAGAGGCCCGGTTCTTCAG									2255
Qy	525	Cys-----Ala	ValPhe	ValVal	AsnGlu	ThrAla	GlyAsn	ProCys	LeuGlu	GluAsn	542
Db	2256	TGCTGGGGCATCGTCAGTTCAGTGTGAGCCGCTCGCTCGCTCGCGCCGTCAGGAGGCC									2315
Qy	543	---Asp	ValVal	LeuSer	PheThr	GluLeu	IleGly	GlyCys	ValPhe	ArgGly	IleArg 561
Db	2316	CTCGATACAAGGATGTTTGTGAGAGATTGTGAGGAAGTGCTGTGAGCTTGGTATCCGT									2375
Qy	562	IleGlyAlaAsn	Glu-----Asn	ArgGly	ValAla	GlnSer	IleMet	TyrAsp	AlaThr		578
Db	2376	ATGAACCCATAATCCATGCTTCGTGCACATAACAAGGATGCGAGTCTCTTCGATCCACAT									2435
Qy	579	LysAsnGlu	TyrAla	PheThr	LysAsn	CysThr	LeuAsn	ThrGly	IleGly	ArgPhe	Glu 598
Db	2436	-----GGACTACATCAAGAGAGCTAAACAAGCAAAA									2465
Qy	599	IleAlaAla	ThrGlu	AlaLys	AsnMet	PheGlu	ArgLeu	ProAsp	LysGlu	GlnLys	Val 618
Db	2466	CAAGCTGCAGTG-----AGCAAGAAGCAGAGGTTG									2495
Qy	619	LeuMetPhe	IleIle	IleSer	LysArg	GlnLeu	AsnAla	TyrGly	PheVal	LysHis	Tyr 638
Db	2496	CAGCTCCTTTCTGCCCCGATGTCGGAGCAGCATTCGGGGTACAAAGACACTGAAGCTGATT									2555
Qy	639	CysAsp	HisThr	IleGly	ValAla	AsnGln	HisIle	ThrSer	GluThr	ValThr	LysAla 658
Db	2556	TGTGACACACAGCTGGGGATCTTGACCGAGTGTTTCTGAGCGACCGCGCA-----									2606
Qy	659	LeuAlaSer	LeuArg	HisGlu	LysGly	SerLys	ArgIle	PheThr	GlnIle	AlaLeu	Lys 678

## RESULT 12

US-11-218-305-19388







1516 ATTGCTAGAAATTTACAGAAAGTTGATGGTATTTGACGACCACTTAAGCTTAAAGCT 1575  
Db  
496 SerValAsnGluGlnIleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeu 515  
QY  
1576 GGGAAATCGTGAAGACATTTTACACGCAATGGTAGTGAACCTTCAACAATAAGAGGCTC 1635  
Db  
516 AsnValValProGluLysGluLeuLysCysAlaValPheValValAsnGluThrAlaGly 535  
QY  
1636 -----ATTAGAGCTTGTAGTGTCTC----- 1653  
Db  
536 AsnProCysLeuGluGluAsnAspValValLysPheTyrThrGluLeuIleGlyGlyCys 555  
QY  
1654 -----GAGAAATCGCGGTGTAAACTTTTCTCACGA-----TGC 1689  
Db  
556 LysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyr 575  
QY  
1690 AATGTCAGGATCTTGTCCGGATCTCATCAAGTGTGGAGGATGAAGGCAATTATGTT 1749  
Db  
576 AspAlaThrLysAsnGluTyrAla---PheTyrLysAsnCysThrLysAsnThrGlyIle 594  
QY  
1750 GATGCT-----CCTTTTGCTGTATTGTAGAGAAATCCTCAATG----- 1788  
Db  
595 GlyArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLys 614  
QY  
1789 ---AGACGGTCACTGCTATAAGAGGGTTGAAGACATGTTTGAACAAGTGAACCTAAG 1845  
Db  
615 -----GluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsnAla 631  
QY  
1846 CTTCTCGGACCAAAAGTTCTTTTGTGTCTTAGCTGAAGGAAGAAATCTGTATTT 1905  
Db  
632 TyrGlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThr 651  
QY  
1906 TATGGCCCTTGAAGAAGAAATGCTTGTGAATTTGGGATCGTTACAAATGTGTGCA 1965  
Db  
652 SerGluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIle 671  
QY  
1966 CCAACTAGAGTG----- 1989  
Db  
672 PheTyrGlnIleAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAsp 691  
QY  
1990 CTTACAAATGTCTACTTAAGATAAAATGCAAAAGCTGGTGGCATGAATCGTTGCTCAA 2049  
Db  
692 TrpSerGluIleAlaGluIleSerProGluLysGluArgLysThrMetProLeu 711  
QY  
2050 -----ATTGAACATCCCCA-----GCAATCTCTCT 2076  
Db  
712 -----ThrMetTyrValGlyIleAspValThrHisProThrSerTyrSer 726  
QY  
2077 GTATCCAAGTCCCAACTATAATCTTGGGAATGATGTCTCACACGGTCTCTCGACAT 2136  
Db  
727 GlyIleAspTyrSerIleAlaValAlaSerIleAsn---ProGlyGlyThrIle 745  
QY  
2137 TCTGATGTACCATCTATTGCTGTGTGTGTAGTCTCTGGAATGGGCTCTTATCTCGAAA 2196  
Db  
746 TyrArgAsnMetIleValThrGlnGluLysCysArgProGlyGluArgAlaValAlaHis 765  
QY  
2197 TACAGAGCTTCTGTCGCGACCACTAACCTAAATGAATGATTCGACTCATTTGTTAAG 2256  
Db  
766 GlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAla 785  
QY  
2257 CCACGGGAAGCTCAAGAT-----GATGCTCTGATCCGGAGTGTCTGATGACTTCTAC 2310  
Db  
786 GluAsnAsnAspAsnArgAlaProAlaHisIleValValTyrArgAspGlyValSerAsp 805  
QY  
2311 ACCAGTTCTCGGAAGAAGACCTGACCAAGTCTCATATTCAGGACGGTGTAGCGAA 2370  
Db  
806 SerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGln 825  
QY  
2371 AGTCAGTTTAATCAGGTGTGAACATTTGATGTGAACAATC---ATCGAGGCTTGCAAA 2427  
Db  
826 PheMetSerGluArgAspGlyGluAspProGluProLysThrPheIleValIleGln 845  
QY  
2428 TTTCTTGATGAGAA-----TGAATCCCAAGTTCACGGTTGATTTATGCCAG 2475  
Db

846 LysArgHisAsnThrArgLeuLeuArgMetGluLysAspLysProValValAsnLys 865  
QY  
2476 AAGAATCATCACACT-----AAA 2493  
Db  
866 AspLeuThrProAlaGluThrAspValAlaValAlaValLysGlnTrpGluGluAsp 885  
QY  
2494 TTTTTCATTCCTGGAAAGCCAGAT----- 2517  
Db  
886 MetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLys 905  
QY  
2518 -----ATGTCCACCAAGAACTGTGTGGACAAAC 2547  
Db  
906 LeuIleValSerLysTyrLysPheAspPheLeuAlaSerHisGlyValLeuGly 925  
QY  
2548 AAAGTCTGCCATCCAAAGAACTTCGATTTCTACATGTGTGGCATGCTGGAATCATCGG 2607  
Db  
926 ThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGlu 945  
QY  
2608 ACTACAGGGCCAACTCACTACCACTCTCATCATGATGATAGTAGGCTTTCAGTCTCATGAT 2667  
Db  
946 ValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSer 965  
QY  
2668 CTCAGAGCTGTGTGATTCGCTCTCTTATGTGTACCAAGGACCAACAGCCATATCA 2727  
Db  
966 LeuProValProValHisTyrAlaHisLeuSerCysGlu-----Lys 979  
QY  
2728 GTGCTTGTCTCCATCTGTACGCACATCTGGCAGCTGCTCAGGTTGGCCAGTTTCATAAAG 2787  
Db  
980 AlaLysGluLeuTyrArgThrTyrLysGluHis 990  
QY  
2788 TTCATGATGATGTCGGAGACGCTCTCCAGTCAT 2820  
Db

## RESULT 13

US-11-218-305-14683  
; Sequence 14683, Application US/11218305  
; Publication No. US2006014195A1  
; GENERAL INFORMATION:  
; APPLICANT: MONSANTO TECHNOLOGY, LLC  
; APPLICANT: McLaird, Paul L.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Wu, Kunsheng  
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
; FILE REFERENCE: 38-21 (53660)B  
; CURRENT APPLICATION NUMBER: US/11/218,305  
; CURRENT FILING DATE: 2005-09-01  
; PRIOR APPLICATION NUMBER: US 60/606,880  
; PRIOR FILING DATE: 2004-09-01  
; NUMBER OF SEQ ID NOS: 25043  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 14683  
; LENGTH: 2012  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-11-218-305-14683

Alignment Scores:  
Pred. No.: 1,17e-37 Length: 2012  
Score: 466.00 Matches: 166  
Percent Similarity: 42.6% Conservatives: 97  
Best Local Similarity: 26.9% Mismatches: 229  
Query Match: 8.7% Indels: 126  
DB: 7 Gaps: 24

US-10-645-746-3 (1-1020) x US-11-218-305-14683 (1-2012)

QY 392 IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTyrAlaVal 411  
Db 27 ATTCAGCATCTCGCAGCAGACCATTTACCTAGAACATCTTAATCGAAAAACGAATATTTA 86  
QY 412 ProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArgIleAspLeu 431  
Db

Db 87 CCAATGGAGCGCTGCAAGATCGTTGAA---GGCCAGAGATACACGAAGAGTTGTAATGAA 143  
Qy 432 ValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLysGluAsn 451  
Db 144 AAACAGATCACATCGTTGCTAAAGGTTACATGCCAAGGCTCGAGAAACAAGAGATGGAT 203  
Qy 452 ThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheValGluArg 471  
Db 204 ATTCTACACAGACATTCATCAAAATGGATAT-----GACCAAGATCCATATCGAAGGAA 257  
Qy 472 PheGly-----LeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys 489  
Db 258 TTTGGGATCAACATTAGTCAGAAGCTAACCTATGTTGAA-----GCCCGAGTCTCTCCT 311  
Qy 490 GluProMetLeu---ValAsnSerValAsnGluGlnIleLysMetThrProValIleArg 508  
Db 312 GCACCTTGGCTGAAGTATCATGACACTGGAAAAGAGAAAGAGTGTCTTACCA-----362  
Qy 509 GlyPheGlnGluLysGlnLeuAsnValValProGluLysGluLeu---CysCysAlaVal 527  
Db 363 -----CAGTTGCTCAGTGGACATGGTAACAAGAAATGTCAAATTTCCGGGCATGGAAATTTAACAGT 527  
Qy 565 AsnGluAsnArgGlyAlaGlnSerIleMetTyrAsp-----AlaThrLysAsn 580  
Db 528 GAGCCGCTGATGCCAATATATTTCAGCTAGACACAGATCAAGTAGTGAAGGCACCTTAAAGT 587  
Qy 581 GluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPheGluIleAla 600  
Db 588 GTGTAT-----AATATTGCACCTGAAC-----608  
Qy 601 AlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLysValLeuMet 620  
Db 609 -----AAACTCAGGGTAAAGAACTTGAACCTTCTTCG 641  
Qy 621 PheIleIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAsp 640  
Db 642 ----GCTATACTCCCGACACAACTATGCTTATATATGTTGATCATCAACAGTATTTGTGAA 698  
Qy 641 HisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLysAlaLeuAla 660  
Db 699 ACTGATTTGGGATTGATATACATATGCTTAAACCAACAGCATGTTTAAAGATC-----752  
Qy 661 SerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLeuLysIleAsn 680  
Db 753 -----AGCAACACAGTACTTGGCAAAATGCTCTCACTGAAATAAT 791  
Qy 681 AlaLysLeuGlyGlyIleAsn-----GlnGluLeuAspTyrSerGluIleAla 696  
Db 792 GTTAGATGGGAGGAGAAACACTGTGCTCGGACGCATTAAGTTGGAGC-----842  
Qy 697 GluIleSerProGluLysGluArgArgLysThrMetProLeu-----711  
Db 843 -----ATTCCTTTGGTCAGTCAGTACATCCCA 866  
Qy 712 ThrMetTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSer 731  
Db 867 ACTATTATATTGGTGCAATGTAACACACCCCTGAAACCGGGAGGAGTCAAGTCCATCA 926  
Qy 732 IleAlaAlaValValAlaSerIleAsn---ProGlyGlyThrIleTyrArgAsnMetIle 750  
Db 927 ATCGCTGCCGTTGCTTCTCAAGATTGGCCAGAGTTACAAAGTATGCTGATGTTGTT 986  
Qy 751 ValThrGln-----GluGluCysArgProGlyGluArgAlaValAlaHisGlyArg 767  
Db 987 TGTGCTCAGGCACACCGGCAAGAGTCTATTTCAGGACCTTTACAAACCATGGCACGATCCT 1046

Qy 768 GluArgThrAspIleLeuGluAlaLysPheValLysPheValLysLeuLeuArgGluPheAlaGluAsn 787  
Db 1047 CAGAGAGGCACCTGTAAACAGCGCGCATGTACAGGAGCTGTTAATATCTTCAGGAAGGCC 1106  
Qy 788 AsnAspAsnArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGlu 807  
Db 1107 ACTGGCCAGAG---CCATTGAGATATATTTCTACAGGACGGTGTATTAGTGAAGCCAG 1163  
Qy 808 MetLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMet 827  
Db 1164 TTCTATCAAGTCTCTCTTTACGAGTTAGATGCCATCCGTAAGGCATCGCATCCCTA---1220  
Qy 828 SerGluArgAspGlyGluAspProGluProLysTyr-----ThrPheIleVal 843  
Db 1221 -----GAACCAAAATTACGACCTCTCTGTAAATTTGGTG 1256  
Qy 844 IleGlnLysArgHisAsnThrArgLeuLeuArgMetGluLysAspLysProValVal 863  
Db 1257 GTTCAAAAACGTCATCATACGAGACTATTTACAAACATCACAACAGACAGTAGCATC 1316  
Qy 864 AsnLysAspLeuThrProAlaGluThrAspValAlaValAlaValLysGlnTrpGlu 883  
Db 1317 GACAAG-----1322  
Qy 884 GluAspMetLysGluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrVal 903  
Db 1323 -----AGTGAATAATTTTGCCA-----GGAAGTGTGT 1352  
Qy 904 AspLysLeuIleValSerLysTyrLysPheAspPhePheLeuAlaSerHisGlyVal 923  
Db 1353 GATTCTAAGATATGCCACCAACAGAGTTGATTTCTACCTCTGTAGTCATGCTGGAATC 1412  
Qy 924 LeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGln 943  
Db 1413 CAGGAAACAAGTAGGCGGCTCACTACCATGTCTCTGGGATGAGAACAAATTTTCACAGCA 1472  
Qy 944 AspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysPro 963  
Db 1473 GACGAATGCAAACTCATCAACAAACCTTTGCTACACTTATGCGCGTGACACGCTCG 1532  
Qy 964 IleSerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAlaLys 981  
Db 1533 GTTCTGTGTCCTCTCTGCATCTACTACGACACCTTGGCAGCATTCGGGCGCGG 1586

## RESULT 14

US-11-218-305-19467  
; Sequence 19467, Application US/11218305  
; Publication No. US20060141495A1  
; GENERAL INFORMATION:  
; APPLICANT: MONSANTO TECHNOLOGY, LLC  
; APPLICANT: McLeird, Paul L.  
; APPLICANT: Tao, Kunsheng  
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
; FILE REFERENCE: 38-21 (53660)B  
; CURRENT APPLICATION NUMBER: US/11/218,305  
; PRIOR FILING DATE: 2005-09-01  
; PRIOR APPLICATION NUMBER: US 60/606,880  
; PRIOR FILING DATE: 2004-09-01  
; NUMBER OF SEQ ID NOS: 25043  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 19467  
; LENGTH: 2773  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-11-218-305-19467

Alignment Scores:  
Pred. No.: 1,976-36 Length: 2773  
Score: 456.00 Matches: 204  
Percent Similarity: 39.9% Conservative: 125

Best Local Similarity: 24.7%  
 Query Match: 8.5%  
 DB: 7  
 Mismatches: 313  
 Indels: 183  
 Gaps: 35

US-10-645-746-3 (1-1020) x US-11-218-305-19467 (1-2773)

QY 225 ValProGluSerPheHisAspProAsnArgPheGluGlnSerLeuGluValAlaProArg 244  
 DB |||||  
 QY 341 GTTGGTAGTCAATTTATCTCCAC---TTAGGAGAGACCCCCCCCCCCCCCCCT 397  
 DB |||||  
 QY 245 IleGluAlaTrpPheGlyIleTyrIleGlyLeuPheAspGlyGluProVal 264  
 DB |||||  
 QY 398 TTGGAAAGTTGGCGGTTTCTACCAAGCATAGGCGGCACACAGATGGGCTTCA--- 454  
 DB |||||  
 QY 265 LeuAsnPheAlaIleValAspLysLeuPheTyrAsn---AlaProLysMetSerLeuLeu 283  
 DB |||||  
 QY 455 CTGAATATTGATATGCTCTACTGCAATTTATCGAGCCTCTCCCTGATTTGTTT 514  
 DB |||||  
 QY 284 AspTyrLeuLeu-----LeuIleValAspProGlnSerCysAsnAspValArg 300  
 DB |||||  
 QY 515 GCTCAGCTTCTTAAGAGATATTTAGTTAGGCATTTGCTGATTTCTGATCGCGTGAAG 574  
 DB |||||  
 QY 301 LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgPro 320  
 DB |||||  
 QY 575 -----ATCAAAAGGCTTAAGAGGTTGAGGTTGAGGCTCACTCACAGGGAAACATG 628  
 DB |||||  
 QY 321 ArgIleArgGlnLeuLeuGluAsnLeuLysCysAlaGluValTrpAspAsnGlu 340  
 DB |||||  
 QY 629 CGCAGAAAGTATCGCATTTCTGCGCTC----- 655  
 DB |||||  
 QY 341 MetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeu 360  
 DB |||||  
 QY 656 ACCTCAACGAACA---AGAGAGCTATCATTC----- 685  
 DB |||||  
 QY 361 ValTyrLysValThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThr 380  
 DB |||||  
 QY 686 -----CTGTTGATGATCGTGGTACT-----GTGAAGACT 715  
 DB |||||  
 QY 381 LeuPheLysIleTyrGluGluAsnLysLysPhe-----IleGluPheProHisLeuProLeu 399  
 DB |||||  
 QY 716 GTGGTCCAACTACTTCATGAGAGCTTATGGTTTATGATCACCACACCACTTTACCGTGC 775  
 DB |||||  
 QY 400 ValLysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHis 419  
 DB |||||  
 QY 776 TTACAAGTGGCAATCAACAAAGACCAATATTATCGCTATGAGGTTTGCAGAGATGTT 835  
 DB |||||  
 QY 420 GluLysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLys 439  
 DB |||||  
 QY 836 GAA---GGACAGCGTTACTCAAGCGCACTCAATGAGAAACAAATCACTGCTCTACTGAAA 892  
 DB |||||  
 QY 440 ArgAlaThrArgLysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeu 459  
 DB |||||  
 QY 893 GTGACCTGCGAGCGCCT-----CAAGACGTCAGCTGAGACATTTTACAGACTGTG 943  
 DB |||||  
 QY 460 AspPheSerSer---GluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeu 478  
 DB |||||  
 QY 944 CATCAATCGCTACTATGAAGACCCGTATGCAAGAAATTTGGTATAAGAAATTGATGAA 1003  
 DB |||||  
 QY 479 GlnMetIleGluCysProGlyLysValLeuLysGluProMetLeu---ValAsnSerVal 497  
 DB |||||  
 QY 1004 CGCCTTGCTGCAAGTCTGAGTCTGTTCTGCCACCAAGGCTTAAATACCATGATAGT 1063  
 DB |||||  
 QY 498 AsnGluGlnIleLysMetThrProValIle-----ArgGlyPheGlnGluLysGln 514  
 DB |||||  
 QY 1064 GGCAGAGAGAGGATGTTTGGCCCAAGATTTGGCAATGGAACATGATGAATAAGAAATG 1123  
 DB |||||  
 QY 515 LeuAsnValValProGluLysGluLeuCysCysAlaValPheValValAsnGluThrAla 534  
 DB |||||  
 QY 1124 GTTAATGGTGGCAGACTCAGCAACTGGCGATGATTAACCTTCTCTCGGAAT----- 1174  
 DB |||||  
 QY 535 GlyAsnProCysLeuGluGluAsnAspValValLysPheTyrThrGluLeuIleGlyGly 554  
 DB |||||  
 QY 1175 -----GTGCAAGTAGTGGCGGTCTGTGTCATGCACTGGCAATCATG 1222  
 DB |||||

QY 555 CysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMet 574  
 DB |||||  
 QY 1223 TGCCAAATATATCAGGAATG-----GATTTTCTCTTGAGCCTGTGCTG 1264  
 DB |||||  
 QY 575 TyrAspAlaThrLysAsn-----GluTyrAlaPheTyrLysAsnCysThrLeu 590  
 DB |||||  
 QY 1265 CTCCAGTGTAGTCGAGGCCAGAACATGTTGGAAGAGCGTTGAAG----- 1309  
 DB |||||  
 QY 591 AsnThrGlyIleGlyArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArg 610  
 DB |||||  
 QY 1310 -----GCAGTTATCAA-----GATGCATGAACATACTAGG--- 1342  
 DB |||||  
 QY 611 LeuProAspLysGluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsn 630  
 DB |||||  
 QY 1343 ---CCACAGGGAGGAACTTGATCTGCTGATTTAATACTGCTGACATTAATGGTTCC 1399  
 DB |||||  
 QY 631 AlaTyrGlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIle 650  
 DB |||||  
 QY 1400 TTATATGGGATCTCAAAAGGATCTGTGACATGTATCGGATTTGGTCTCCCATGTTGT 1459  
 DB |||||  
 QY 651 ThrSerGluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArg 670  
 DB |||||  
 QY 1460 CTGACTAAACATGTTTTTAAGATG-----AGCAGCAG 1492  
 DB |||||  
 QY 671 IlePheTyrGlnIleAlaLeuLysIleAsnAlaLysLeuGly---GlyIleAsnGlnGlu 689  
 DB |||||  
 QY 1493 TATCTTGCAAAATGTTGCACTCAAAATAAATGTTAAGTTGCGCGGCGCGCCCT 1552  
 DB |||||  
 QY 690 LeuAspTrpSerGluIleAlaGluIleSerProGluGluLysGluArgArgLysThrMet 709  
 DB |||||  
 QY 1553 TTAGAT-----GCTTTGACAAGAGGAATC 1576  
 DB |||||  
 QY 710 ProLeu-----ThrMetTyrValGlyIleAspValThrHisProThrSer 724  
 DB |||||  
 QY 1577 CCCTTTGTCAGTGACAGACCGACCAATAATTTGCTGCTGATGTTACCATCCATCCT 1636  
 DB |||||  
 QY 725 TyrSerGlyIleAspTyrSerIleAlaValAlaSerIleAsn-----ProGlyGly 743  
 DB |||||  
 QY 1637 GGAGAAAGATTCACAGTCTCTCCATTGCACTGTGTTGCTTTCGCAAGACTGCTGAGTC 1696  
 DB |||||  
 QY 744 ThrIleTyrArgAsnMetIleValThrGlnGluCysArgProGlyGluArgAlaVal 763  
 DB |||||  
 QY 1697 ACCAAATATGCTCGACTAGTGTAGTGCCTCA----- 1726  
 DB |||||  
 QY 764 AlaHisGlyArgGlu-----ArgThr 770  
 DB |||||  
 QY 1727 GCCATCGCAGGAGCTGATACAGGATCTTTCAAAGTATGGCAAGATCCACAGAGAAG 1786  
 DB |||||  
 QY 771 AspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsn 790  
 DB |||||  
 QY 1787 ACAGTAACCTGGTGGCATGATAAGGAATCTCTCATTTCTTTC---AAGAGAGCAACTGGA 1843  
 DB |||||  
 QY 791 ArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArg 810  
 DB |||||  
 QY 1844 CAGAAGCCCCAGAGGATCATATTTCTACAGGATGGTGTGAGTGAGACAGTCTTATCAA 1903  
 DB |||||  
 QY 811 ValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArg 830  
 DB |||||  
 QY 1904 GTATTGTTGATGAATTCATGCCATCAGAAAGGCTGTGTCATCTTG----- 1951  
 DB |||||  
 QY 831 AspGlyGluAspProGluProLysTyr-----ThrPheIleValIleGlnLys 846  
 DB |||||  
 QY 1952 -----GAGCCCAACTACCAAGCTCCAGTCTTACTTTTCTGCTGGTGCAGAAA 1996  
 DB |||||  
 QY 847 ArgHisAsnThrArgLeuLeuArgArgMetGluLysAspLysProValValAsnLysAsp 866  
 DB |||||  
 QY 1997 CGCATCATCAGTGGCTGTTTGTCTAATAACCAACAGCATCAGGTACAGTTGATAGA--- 2053  
 DB |||||  
 QY 867 LeuThrProAlaGluThrAspValAlaValAlaValLysGlnTrpGluGluAspMet 886  
 DB |||||  
 QY 2053 ----- 2053  
 DB |||||

QY 887 LysGluSerLysGluThrGlyIleValAenProSerSerGlyThrValAspLysLeu 906  
 Db 2054 -----AGCGGAACATACATCGCCT-----GGCACCGTGTGATTGGAAG 2092  
 QY 907 IleValSerLysThrLysPhePheLeuAlaSerHisGlyValLeuGlyThr 926  
 Db 2093 ATTGCCATCTACTGAATTTGATTCTACTGTGTAGCCATCTGGCATTGAGGAC 2152  
 QY 927 SerArgProGlyHisThrValMetTyrAspLysGlyMetSerGlnAspGluVal 946  
 Db 2153 AGCGCCCTGCTCAATTACCATCTCTGTGGGACGAGAACAGTTCCACAGCTGATAGCTG 2212  
 QY 947 TyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaAtqCysArgLysProIleSerLeu 966  
 Db 2213 CAGACTCTGACAAACACTATGCTACACGTACGTAGGTGACCGCTCCGTGCAATT 2272  
 QY 967 ProValProValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThr 986  
 Db 2273 GTGCCCGGCATCATCTGCTGGCGGCTTCCGAGCT-----CGCTTC 2320  
 QY 987 TyrLysGluHisTyrIleGlyAsp-TyrAlaGln-----ProArgThrArgHis 1002  
 Db 2321 TACATGGACCCAGATACCTCTGACAGTGGCTCAATGGCCAGTGGTGGCCGCTCCCA 2380  
 QY 1002 sGluMetGluHis 1006  
 Db 2381 CCAGGTGGCGCAC 2393

## RESULT 15

US-11-218-305-19466  
 ; Sequence 19466, Application US/11218305  
 ; Publication No. US2006014195A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MONSANTO TECHNOLOGY, LLC  
 ; APPLICANT: McLaird, Paul L.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Wu, Kunheng  
 ; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
 ; TITLE OF INVENTION: Corn.  
 ; FILE REFERENCE: 38-21 (53660)B  
 ; CURRENT APPLICATION NUMBER: US/11/218,305  
 ; CURRENT FILING DATE: 2005-09-01  
 ; PRIOR APPLICATION NUMBER: US 60/606,880  
 ; PRIOR FILING DATE: 2004-09-01  
 ; NUMBER OF SEQ ID NOS: 25043  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 19466  
 ; LENGTH: 3827  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (480)..(480)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (515)..(515)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (568)..(568)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (649)..(649)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (787)..(787)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (964)..(970)

OTHER INFORMATION: n is a, c, g, or t  
 US-11-218-305-19466

## Alignment Scores:

Pred. No.: 3,97e-36 Length: 3827  
 Score: 455.00 Matches: 196  
 Percent Similarity: 40.6% Conservative: 126  
 Best Local Similarity: 24.7% Mismatches: 301  
 Query Match: 8.5% Indels: 173  
 DB: 7 Gaps: 32

US-10-645-746-3 (1-1020) x US-11-218-305-19466 (1-3827)

QY 225 ValProGluSerPheHisAspProAenArgPheGluGlnSerLeuGluValAlaProArg 244  
 Db 1388 GTCGGTAGTCACTTTTACTCTCCCAAC---TTAGGGAGAGCGTCAAAAACCTTGGTAGGGA 1444  
 QY 245 IleGluAlaTTPheGlyIleTyrIleGlyIleLysGluLeuPheAspGlyGluProVal 264  
 Db 1445 TTGGAAAGTTGGCGTGGTTTTTACCAAGCATAGCCGACACA-GATGGGCTTT-TCA 1502  
 QY 265 LeuAenPheAlaIleValAspLysLeuPheTyrAen---AlaProLysMetSerLeuLeu 283  
 Db 1503 CTGAATATTGATATGCTCTACTTCATTTATCGAGCCTCTCCTGTGATCGATTGTGT 1562  
 QY 284 AspTyrLeuLeu-----LeuIleValAspProGlnSerCysAenAspAspValArg 300  
 Db 1563 GCTCAGCTTCTTAACAGAGATATCTCAGTAGGCCATTTGCTGTGATCTGTCGCGTGAAG 1622  
 QY 301 LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgPro 320  
 Db 1623 -----ATTAAAAAGCCCTAAGAGGTGTGAAGTTGAGGTGACTCAGACGGGAAACATG 1676  
 QY 321 ArgIleArgGlnLeuLeuGluAenLysLysLeuLysCysAlaGluValTrpAspAenGlu 340  
 Db 1677 CGCAAAAATATCGCATTTCTGGCCCTC-----1703  
 QY 341 MetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeuLysGluGluAenSerLeu 360  
 Db 1704 ACCTCACAGCAACA---AGAGAGCTATCATTC-----1733  
 QY 361 ValTyrLysValThrGlyLysSerAspArgGlyArgAenAlaLysLysTyrAspThrThr 380  
 Db 1734 -----CCTGTTGATGATCGTGTACT-----GTGAAGACT 1763  
 QY 381 LeuPheLysIleTyrGluGluAenLysLysPhe-----IleGluPheProHisLeuProLeu 399  
 Db 1764 GTGGTGCAATACTTTCATGGAGACTTATGGTTTTTAGTATCCACACACCATCTTACCATGC 1823  
 QY 400 VallysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluHis 419  
 Db 1824 TTGCAAGTGGGTAATCAACAAGACCAAAATATCTGCCCTATGGAGGTTTGCAGATAGATT 1883  
 QY 420 GluLysProGlnArgTyrLysAenArgIleAspLeuValMetGlnAspLysPheLeuLys 439  
 Db 1884 GAA---GGCAGCGTTACTCAAAGCGACTCAATGAGAAACAATCACTCTCTACTGAAA 1940  
 QY 440 ArgAlaThrArgLysProHisAspTyrLysGluAenThrLeuLysMetLeuLysGluLeu 459  
 Db 1941 GTGACCTGCCAGCGCCCT-----CAAGAGCGGAGCTGGACATCTTACAGACTGTG 1991  
 QY 460 AspPheSerSer---GluGluLeuAenPheValGluArgPheGlyLeuCysSerLysLeu 478  
 Db 1992 CATCAATGATCACTATGTAAGACCCCTATGCAGCTGGAAATTTGGTATAGAAATGATGAA 2051  
 QY 479 GlnMetIleGluCysProGlyLysValLeuLysGluProMetLeu---ValAenSerVal 497  
 Db 2052 CGTCTGTGTCAGTTGAAGCTCGTGTCTGCCACCACCAAGACTTAAATACCATGATAGT 2111  
 QY 498 AenGluGlnIleLysMetThrProValIle-----ArgGlyPheGlnGluLysGln 514  
 Db 2112 GCGCCGAGAGAGGATGTTTGGCCCACTGGCCCAATGGAACATGATGAATAAGAAATG 2171

```
Qy 515 LeuAsnValValProGluLysGluLeuCysCysAlaValPheValValAsnGluThrAla 534
Db 2172 GTTAAATGTCGACAGTGCAGCAACTGGGCATGTATTAACTTCTCTCGGAAT----- 2222
Qy 535 GlyAsnProCysLeuGluGluAsnAspValValLysPheTyrThrGluLeuIleGlyGly 554
Db 2223 -----GTGCAAGATAGTGGCCCTAGGCGTTCCTCATGAGTTGGCAGTCATG 2270
Qy 555 CysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMet 574
Db 2271 TGCCAAAATATCAGGAATG-----GATTGTGCTTTCAGCCGTGCTG 2312
Qy 575 TyrAspAlaThrLysAsn-----GluTyrAlaPheTyrLysAsnCysThrLeu 590
Db 2313 CCTCAGTACTGCAAGCCAGAACATGTTGAGAGAGCGTTAAAG----- 2357
Qy 591 AsnThrGlyIleGlyArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArg 610
Db 2358 -----GCACGTTATCAA-----GATGCAATGAACATACTAGG--- 2390
Qy 611 LeuProAspLysGluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsn 630
Db 2391 ---CCACAGGAAGGAACCTGATCTGCTGATCGTAATACTGCTGACCAACAATGGTTCT 2447
Qy 631 AlaTyrGlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIle 650
Db 2448 CTTTATGGGATCTCAAAAGGATCTGTGAGACTGAATCGGATGGTCTCCCACTGTTGT 2507
Qy 651 ThrSerGluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArg 670
Db 2508 CTGACTAAACATGTTTAAAGT-----AGGCAACAG 2540
Qy 671 IlePheTyrGlnIleAlaLeuLysIleAsnAlaLysLeuGlyIleAsnGlnGluLeu 690
Db 2541 TACCTTGCAAAATGTTGCACTCAAAATAATGTTAAAGTGGGGAAGGAATACTGTACTT 2600
Qy 691 AspTrpSerGluIleAlaGluIleSerProGluLysGluArgGlySerThrMetPro 710
Db 2601 -----TTAGATGCTTTGTCA-----AGGAGAATCCCC 2627
Qy 711 Leu-----ThrMetTyrValGlyIleAspValThrHisProThrSerTyr 725
Db 2628 CTTGTCAGTCAGACAGCCGACCAATAATTTGGTCTGATTTACCCATCCATCCTCGGA 2687
Qy 726 SerGlyIleAspTyrSerIleAlaAlaValAlaSerIleAsn---ProGlyGlyThr 744
Db 2688 GAAGATTCCAGTCTTCCATTGACGCGTGTGTTCTTCGCAAGACTGGCCCGGACGTCACG 2747
Qy 745 IleTyrArgAsnMetIleValThrGlnGluCysArgProGlyGluArgAlaValAla 764
Db 2748 AATACGCTCGACTAGTAGTGCGCAA-----GCC 2777
Qy 765 HisGlyArgGlu-----ArgThrAsp 771
Db 2778 CATCGCCAGAGCTGATACAGGATCTTTCAAAGTATGGCAGGCCGCGAGAGAGGACG 2837
Qy 772 IleLeuGluAlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsnArg 791
Db 2838 GTAACCTGGCGCATATAAAGGAACCTCTCATTTCTTC---AAGAGGGCAACTGGACAG 2894
Qy 792 AlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgVal 811
Db 2895 AAGCCCCAGAGGATCATATTCTACAGGATGGTGTCTAGTGGGACAGATTCTATCAAGTA 2954
Qy 812 SerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAsp 831
Db 2955 TTGCTGTACGAACCTTGATGTCATTAGAACGCCCTGTGCTCCCTG----- 2999
Qy 832 GlyGluAspProGluProLysTyr-----ThrPheIleValIleGlnLysArg 847
Db 3000 -----GAGCCCAACTACCAGCCTCCAGTTACTTTTGTGTTGTTACAGAGCGC 3047
Qy 848 HisAsnThrArgLeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeu 867
```

```
Db 3048 CATCACATAGGCTGTTTGGAAACAACACAGTATCAGCGCACAGTCGATAGA----- 3101
Qy 868 ThrProAlaGluThrAspValAlaValAlaAlaValLysGlnTrpGluGluAspMetLys 887
Db 3101 ----- 3101
Qy 888 GluSerLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeuIle 907
Db 3102 -----AGCGGAAACATACTGCT-----GGCACCGTGGTTCGATTTCGAAGATT 3143
Qy 908 ValSerLysTyrLysPheAspPhePheLeuAlaSerHisHisGlyValLeuGlyThrSer 927
Db 3144 TGCCATCCTACTGAGTTTGACTTCTTACTGTGTAGCCATGCTGCATTTACGGGAACGAGC 3203
Qy 928 ArgProGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyr 947
Db 3204 CGCCCTGCTCACTACCATGCTGCTGTGGAGCGAGAACAAAGTTACACAGCTGACGAGCTGCAG 3263
Qy 948 LysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuPro 967
Db 3264 ACCCTGACGAACAACCTGTGTACACGTCACGTCAGTAGGTGCACCCGCTCCGTGTCCATCGT 3323
Qy 968 ValProValHisTyrAlaHisLeuSerCysGluLysAlaLys 981
Db 3324 CCCCCGCGTACTACGCTCATCTGGCAGCCCTTCCGAGCTCGC 3365
```

Search completed: July 6, 2006, 07:16:19  
Job time : 1546 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2006, 21:39:41 ; Search time 14570 Seconds  
(without alignments)  
12385.153 Million cell updates/sec

Title: US-10-645-746-2

Perfect score: 3227  
Sequence: 1 cagccacaagtgatgaac.....tttaaaaaaaaaaaaaaa 3227

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est3: \*  
3: gb\_est4: \*  
4: gb\_est5: \*  
5: gb\_est6: \*  
6: gb\_hic: \*  
7: gb\_est2: \*  
8: gb\_est7: \*  
9: gb\_est8: \*  
10: gb\_est9: \*  
11: gb\_gss1: \*  
12: gb\_gss2: \*  
13: gb\_gss3: \*  
14: gb\_gss4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	782.4	24.2	796	2	BJ155598
C 2	774	24.0	786	2	BJ143897
C 3	697.4	21.6	700	2	BJ150328
C 4	695.8	21.6	710	2	BJ132853
C 5	689	21.4	700	2	BJ155451
C 6	677	21.0	677	2	BJ779221
C 7	629.8	19.5	633	2	BJ122122
C 8	603.4	18.7	635	2	BJ755060
C 9	582	18.0	583	2	BJ116268
C 10	563.4	17.5	565	4	CB401772
C 11	555.4	17.2	557	2	BJ105221
C 12	554	17.2	554	4	CB398209
C 13	534	16.5	534	2	BJ127233
C 14	527	16.3	539	2	BJ127087
C 15	416.4	12.9	449	9	D27223
C 16	393.8	12.2	421	9	D27221
C 17	375	11.6	378	4	C62850
C 18	370.8	11.5	375	9	D32492
C 19	359	11.1	360	4	C60407

20	358	11.1	360	9	D35146
21	356.4	11.0	360	9	D35838
22	354	11.0	356	10	Z14900
23	348	10.8	360	4	C65335
24	345.4	10.7	370	1	AV191613
25	339.8	10.5	345	9	D33093
26	339.6	10.5	373	4	C60787
27	331	10.3	376	4	C60885
28	309	9.6	340	9	D27220
29	293.4	9.1	334	10	M89235
30	289.8	9.0	300	4	C51768
31	288	8.9	300	4	C51841
32	285	8.8	300	4	C51449
33	282.2	8.7	300	4	C53264
34	276.4	8.6	295	9	D27222
35	262	8.1	300	4	C54925
36	77	2.4	701	6	AV809756
37	77	2.4	701	9	CX861295
38	77	2.4	715	3	BU723459
39	71.8	2.2	834	13	CZ532832
40	69.2	2.1	749	8	CV511740
41	69.2	2.1	773	8	CV508440
42	69	2.1	577	2	BJ113168
43	69	2.1	623	9	DN303856
44	68.8	2.1	667	1	AU217048
45	68.8	2.1	703	2	BJ815785

## ALIGNMENTS

RESULT 1  
BJ155598/c

LOCUS

DEFINITION BJ155598 unpublished oligo-capped cDNA library, C. elegans L1 stage  
Caenorhabditis elegans cDNA clone yk1344b04 3', mRNA sequence.

ACCESSION BJ155598

VERSION BJ155598.1

KEYWORDS EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

796 bp mRNA linear EST 24-JAN-2002

Caenorhabditis elegans cDNA library, C. elegans L1 stage

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 796)

Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

and Sugano, S.

A complementary view of the C. elegans genome

Unpublished (2002)

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1..796

/organism="Caenorhabditis elegans"

/mol\_type="mRNA"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk1344b04"

/sex="hermaphrodite"

/tissue type="whole animal"

/dev stage="L1"

/clone\_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"

ORIGIN

Query Match

Best Local Similarity

Matches 794; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

24.2%; Score 782.4; DB 2; Length 796;

Best Local Similarity 99.6%; Pred. No 4e-174;

Matches 794; Conservative 0; Mismatches 0; Indels 1; Gaps 1;





Qy	3134	TTGTACTTAC	TTTTTTTAAAGCCGGTTTCAAAAATTCATTCATGACTAAACGTTTTCA	3193
Db	66	TTGTACTTAC	TTTTTTTAAAGCCGGTTTCAAAAATTCATTCCTCA-GACTAAACGTTTTCA	8
Qy	3194	TAAATTA	3200	
Db	7	TAAATTA	1	
RESULT 3				
BIJ50328/c				
LOCUS				
DEFINITION	BIJ50328	700 bp	mRNA	linear
ACCESSION	BIJ50328	unpublished	oligo-capped	cDNA library, C. elegans
VERSION	BIJ50328	1		
KEYWORDS	EST.			
SOURCE	BIJ50328.1	GI:18318313		
ORGANISM	Caenorhabditis elegans			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.			
AUTHORS	1 (bases 1 to 700)			
TITLE	Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y., and Sugano,S.			
JOURNAL	A complementary view of the C.elegans genome			
COMMENT	Unpublished (2002)			
FEATURES	source			
	1..700			
	Location/Qualifiers			
	/organism="Caenorhabditis elegans"			
	/mol_type="mRNA"			
	/strain="N2"			
	/db_xref="taxon:6239"			
	/clone="yk1284h06"			
	/sex="hermaphrodite"			
	/tissue_type="whole animal"			
	/dev_stage="L1"			
	/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"			
ORIGIN				
Query Match	21.6%	Score	697.4;	DB 2; Length 700;
Best Local Similarity	99.9%	Pred. No.	5.5e-154;	
Matches	698; Conservative	0; Mismatches	1; Indels	0; Gaps 0;
Qy	2480	AAGCGAAGTAAACAAATTCATGTGCGAACCGGATGGAGAAGATCCAGAGCCGAAGTACAC	2539	
Db	700	AAGCGAAGTAAACAAATTCATGTGCGAACCGGATGGAGAAGATCCAGAGCCGAAGTACAC	641	
Qy	2540	GTTCAATGTGATTCAGAAAAGACACATACCAATGCTTCGAGAGATGAAAAAGATAA	2599	
Db	640	GTTCAATGTGATTCAGAAAAGACACATACCAATGCTTCGAGAGATGAAAAAGATAA	581	
Qy	2600	GCCAGTGGTCAATAAAGATCTTACTCTGCTGAAACAGATGTCGCTGTTCTGCTGTAA	2659	
Db	580	GCCAGTGGTCAATAAAGATCTTACTCTGCTGAAACAGATGTCGCTGTTCTGTTAA	521	
Qy	2660	ACAATGGAGGAGGATATGAAGAAGCAAGAAACCTGGAATTTGTGAACCCATCATCCGG	2719	
Db	520	ACAATGGAGGAGGATATGAAGAAGCAAGAAACCTGGAATTTGTGAACCCATCATCCGG	461	
Qy	2720	AACAACGTGCGATAAACCTTATCGTTTCGAATAACAATTCGATTTTCTTTGCGACTCA	2779	
Db	460	AACAACGTGCGATAAACCTTATCGTTTCGAATAACAATTCGATTTTCTTTGCGACTCA	401	
Qy	2780	TCATGGTGTCTTGGTATCATCTCGTCAGGACATTTACATCTGTATGTATGACCATTAAGG	2839	



```

Db      77  TCAAAATTTTCAAATGTAGATATTGTACTACTTTTAAAAAGCCGGTTTCAAAAATT 18
Qy      3172  CATTCCATGACTAACGT 3188
Db      17  CATTCCATGACTAACGT 1

RESULT 7
Bj121212
LOCUS      Bj121212                633 bp      mRNA      linear      EST 23-JAN-2002
DEFINITION      Bj121212 unpublished oligo-capped cDNA library, C. elegans L1 stage
                  Caenorhabditis elegans cDNA clone yk1284h06 5', mRNA sequence.
ACCESSION      Bj121212
VERSION      Bj121212.1 GI:18282260
KEYWORDS      EST.
SOURCE      Caenorhabditis elegans
ORGANISM      Caenorhabditis elegans
              Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
              Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE      1 (bases 1 to 633)
AUTHORS      Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.,
              and Sugano,S.
TITLE      A complementary view of the C.elegans genome
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tehini@genes.nig.ac.jp.
              Location/Qualifiers
FEATURES             source
                     1..633
                     /organism="Caenorhabditis elegans"
                     /mol_type="mRNA"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1284h06"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"

ORIGIN

Query Match      19.5%; Score 629.8; DB 2; Length 633;
Best Local Similarity 99.7%; Pred. No. 5.7e-138;
Matches 631; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1377  CTA AAAATGCTGAAGAATTTGGATTCTCTCTTCGAGAGCTAAATTTCTGTGAAGATTTT 1436
Db      1  CTA AAAATGCTGAAGAATTTGGATTCTCTCTTCGAGAGCTAAATTTCTGTGAAGATTTT 60

Qy      1437  GGA TTA TGCTCCAAA CTTCAGATGATCGAATGCCAGGAAAAGT TTTGAAAGAGCCAA TG 1496
Db      61  GGA TTA TGCTCCAAA CTTCAGATGATCGAATGCCAGGAAAAGT TTTGAAAGAGCCAA TG 120

Qy      1497  CTTGTGAATAGTGTAATGAACAAATTA AAAATGACACCA CGATGATCGTGGATTTCAAGAA 1556
Db      121  CTTGTGAATAGTGTAATGAACAAATTA AAAATGACACCA CGATGATCGTGGATTTCAAGAA 180

Qy      1557  AAACAATTGAATGTGTTCCCGAAAAGAACTTTGCTGTGCTGTTTGTAGTCAACGAA 1616
Db      181  AAACAATTGAATGTGTTCCCGAAAAGAACTTTGCTGTGCTGTTTGTAGTCAACGAA 240

Qy      1617  ACAGCGGGAAATCCATGCTTTAGAAGAGAACGACGTTGTTTAAGTCTTACACCGCAACTTA 1676
Db      241  ACAGCGGGAAATCCATGCTTTAGAAGAGAACGACGTTGTTTAAGTCTTACACCGCAACTTA 300

Qy      1677  GGTGGTTGCAAGTTC CGTGGAAATACGAATTTGGTGGCCAATGAAAAACAGAGAGCGCAATCT 1736
Db      301  GGTGGTTGCAAGTTC CGTGGAAATACGAATTTGGTGGCCAATGAAAAACAGAGAGCGCAATCT 360

```

```

QY 1737 ATTATGTACGACGCGACGAAATAATGATGCTTCTTACAAAATTTGTACATAATACC 1796
DB 361 ATTATGTACGACGCGACGAAATAATGATGCTTCTTACAAAATTTGTACATAATACC 420
QY 1797 GGAATCGGTAGATTTCGAAATAGCCGCAACAGAGCGAAGAAATATGTTTGAACGTCTTCCC 1856
DB 421 GGAATCGGTAGATTTCGAAATAGCCGCAACAGAGCGAAGAAATATGTTTGAACGTCTTCCC 480
QY 1857 GATAAGAAACAAAAGCTTTAATGTTCAATATCATTTTCCAAACGCAACTGAAATGCTTAC 1916
DB 481 GATAAGAAACAAAGCTTTAATGTTCAATATCATTTTCCAAACGCAACTGAAATGCTTAC 540
QY 1917 GGTTTTGTGAACATTATTTGCGATCACACATCGGTGTAGCTTAATCAGCATATTAATCTT 1976
DB 541 GGTTTTGTGAACATTATTTGCGATCACACATCGGTGTAGCTTAATCAGCATATTAATCTT 600
QY 1977 GAAACAGTCACAAAAGCTTTGGCATCACTAAGG 2009
DB 601 GAAACAGTCACAAAAGCTTTGGCATCACTAAGG 633

RESULT 8
LOCUS BJ755060 635 bp mRNA linear EST 24-MAY-2004
DEFINITION BJ755060 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone yk1409d07 5', mRNA sequence.
ACCESSION BJ755060
VERSION BJ755060.1 GI:47594822
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 635)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadaeu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .635
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/db_xref="taxon:6239"
/clone="yk1409d07"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="unpublished oligo-capped cDNA library"

FEATURES
source
Query Match 18.7%; Score 603.4; DB 2; Length 635;
Best Local Similarity 99.8%; Pred. No. 1e-131; 1; Indels 0; Gaps 0;
Matches 604; Conservative 0; Mismatches 1;

QY 1 CAGCCCAAAAGTGATGAACAATGCTCTCGAATTTTCCGAAATTTGGAAAAGGATTTATC 60
DB 31 CAGCCCAAAAGTGATGAACAATGCTCTCGAATTTTCCGAAATTTGGAAAAGGATTTATC 90
QY 61 GTCATTCTCGATCCGAGATGAATGCTTCGAGGCCCACTGTTAAATCGACGGCA 120
DB 91 GTCATTCTCGATCCGAGATGAATGCTTCGAGGCCCACTGTTAAATCGACGGCA 150
QY 121 AATTCTTATGAGAAGAAAGTACTTCTTTTGGTAAATGGTTTCAAGTTCTCCAGCAAAATTT 180

```

```

DB 151 AATTCCTATGAGAAAGTACTTCTTTGGTAAATGGTTTCAAGTTCTCCAGCAAAATTT 210
QY 181 ACGATCGGGAATACCTACGAGTATGAAGTGAATAATGACAAAGGAAGTATTTGAATAGAAAAC 240
DB 211 ACGATCGGGAATACCTACGAGTATGAAGTGAATAATGACAAAGGAAGTATTTGAATAGAAAAC 270
QY 241 CAGGAAAACCTTTCCCAAAAAGACAGAAATTTCCAATTTCCCGATCGTGCAAAACTCTTCT 300
DB 271 CAGGAAAACCTTTCCCAAAAAGACAGAAATTTCCAATTTCCCGATCGTGCAAAACTCTTCT 330
QY 301 GCACAATCTTCGGCATGAGAAGAGCAGACAGATTTTATTCTCGAAGACTATGTTTTTG 360
DB 331 GGTACATCTTCGGCATGAGAAGAGCAGACAGATTTTATTCTCGAAGACTATGTTTTTG 390
QY 361 ATGAAAAGGACACTGTTTATAGTGTTCGACTGAACTGACATCAAAATGCTGG 420
DB 391 ATGAAAAGGACACTGTTTATAGTGTTCGACTGAACTGACATCAAAATGCTGG 450
QY 421 TTTTCGGAGAAAGTAGTAAAAAAGGATTCGGAGAAAAGAGATGAAAGAGGATTTGGAGAAA 480
DB 451 TTTTCGGAGAAAGTAGTAAAAAAGGATTCGGAGAAAAGAGATGAAAGAGGATTTGGAGAAA 510
QY 481 AATCTTATACAAATGATCTTACTATCGTAAAAAATTTTCCCTGAACTTTAGTCGAG 540
DB 511 AATCTTATACAAATGATCTTACTATCGTAAAAAATTTTCCCTGAACTTTAGTCGAG 570
QY 541 AAAATCCGAAAAAGACGAAAGCGAATCGGAGTTACAAATTCCTGAAGATGTTATGA 600
DB 571 AAAATCCGAAAAAGACGAAAGCGAATCGGAGTTACAAATTCCTGAAGATGTTATGA 630
QY 601 CCCAG 605
DB 631 CCCAG 635

RESULT 9
LOCUS BJ116268 583 bp mRNA linear EST 23-JAN-2002
DEFINITION BJ116268 unpublished oligo-capped cDNA library, C. elegans L1 stage
Caenorhabditis elegans cDNA clone yk1201c08 5', mRNA sequence.
ACCESSION BJ116268
VERSION BJ116268.1 GI:18276379
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 583)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadaeu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .583
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/db_xref="taxon:6239"
/clone="yk1201c08"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"

FEATURES
source

```

```
Email: Marc.Vidal@fci.harvard.edu
Sequence tag of Gateway entry clones. The primers used were
designed on the predicted protein encoding ORF. C. elegans ORFome
cloning project : Contact david_hill@fci.harvard.edu or
marc_vidal@fci.harvard.edu
POLYA=No. Location/Qualifiers
1..565
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"
```

<b>RESULT 11</b>					
BJ105221					
LOCUS	BJ105221	557 bp	mRNA	linear	EST 18-JAN-2002
DEFINITION	BJ105221 unpublished oligo-capped cDNA library, <i>C. elegans</i> l1 stage Caenorhabditis elegans cDNA clone yk1070fi2 5', mRNA sequence.				



```
ACCESSION BJ105221
VERSION BJ105221.1 GI:18247891
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 557)
AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..557
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1070f12"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cdna library, C.
elegans L1 stage"
ORIGIN
Query Match 17.2%; Score 555.4; DB 2; Length 557;
Best Local Similarity 99.8%; Pred. No. 2.4e-120;
Matches 556; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2043 ATTGCATTGGAATCAACCGGAATTTAGGAGTATTAAACGAGGCTTGACTGGTCAGAA 2102
DB 1 ATTGCATTGGAATCAACCGGAATTTAGGAGTATTAAACGAGGCTTGACTGGTCAGAA 60
QY 2103 ATTGCAGAATATCACCAGAAGAAAAAGAAAGACGCGAAACCAATGCCATTAACTATGTAT 2162
DB 61 ATTGCAGAATATCACCAGAAGAAAAAGAAAGACGCGAAACCAATGCCATTAACTATGTAT 120
QY 2163 GTTGGAAATTGATGTAATCATCAACCTCTACAGTGGAAATTGATTTCTATAGCGGCT 2222
DB 121 GTTGGAAATTGATGTAATCATCAACCTCTACAGTGGAAATTGATTTCTATAGCGGCT 180
QY 2223 GTAGTAGCGGATATCAATCCAGGTGGAATATCTATCGAAATATGATTGACTCAAGAA 2282
DB 181 GTAGTAGCGGATATCAATCCAGGTGGAATATCTATCGAAATATGATTGACTCAAGAA 240
QY 2283 GAATGTGTCGCCGTGAGCGGTGAGTGGCTCATGGACGGGAAGAAACAGATATTTTGGAA 2342
DB 241 GAATGTGTCGCCGTGAGCGGTGAGTGGCTCATGGACGGGAAGAAACAGATATTTTGGAA 300
QY 2343 GCMAAGTTCGTGAAATTGCTCAGAGAAATTCGCGGAAAAACACGACAAATCAGCACCAGCG 2402
DB 301 GCMAAGTTCGTGAAATTGCTCAGAGAAATTCGCGGAAAAACACGACAAATCAGCACCAGCG 360
QY 2403 CATATTGTAGTCTATCGACGCGGATAGCGATTCGGAGATGCTACGTGTTAGTCATGAT 2462
DB 361 CATATTGTAGTCTATCGACGCGGATAGCGATTCGGAGATGCTACGTGTTAGTCATGAT 420
QY 2463 GAGCTTCGATCTTTAAAAAGCGAAGTAAACCAATTCATGTCCGAAACGGGATGAGAGAAT 2522
DB 421 GAGCTTCGATCTTTAAAGAGCGAAGTAAACCAATTCATGTCCGAAACGGGATGAGAGAAT 480
QY 2523 CCAGAGCCGAAGTACACGTTTCATTGTGATTCAGAAAGACACATACAGATGCTTCGA 2582
DB 481 CCAGAGCCGAAGTACACGTTTCATTGTGATTCAGAAAGACACATACAGATGCTTCGA 540
QY 2583 AGAATGGAAAAAGATAA 2599
DB 541 AGAATGGAAAAAGATAA 557
RESULT 12
LOCUS CB398209/c
DEFINITION CB398209
ACCESSION CB398209.1 GI:30739936
VERSION CB398209.1
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 554)
AUTHORS Reboul,J., Vaglio,P., Rual,J.F., Lamesch,P., Martinez,M.,
Armstrong,C.M., Li,S., Jacotot,L., Bortin,N., Janky,R., Moore,T.,
Hudson,J.R., Hartley,J.L., Braach,M.A., Vandenhaute,J., Boulton,S.,
Endress,G.A., Jena,S., Chevet,E., Papasotiropoulos,V.,
Tollas,P.P., Placek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H.,
Doucette-Stamm,L., Hill,D.E. and Vidal,M.
TITLE C. elegans ORFome version 1.1: experimental verification of the
genome annotation and resource for proteome-scale protein
expression
JOURNAL Nat. Genet. (2003) In press
COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Sequence tag of Gateway entry clones. The primers used were
designed on the predicted protein encoding ORF. C. elegans ORFome
cloning project : Contact david_hille@dfci.harvard.edu or
marc_vidal@dfci.harvard.edu
POLYA=No. Location/Qualifiers
1..554
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A) +
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"
ORIGIN
Query Match 17.2%; Score 554; DB 4; Length 554;
Best Local Similarity 100.0%; Pred. No. 5.2e-120;
Matches 554; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2510 GGATGGAGAAGATCCAGAGCCGAAAGTACACGTTTCATTGTGATTCAGAAAAAGACAAATAC 2569
DB 554 GGATGGAGAAGATCCAGAGCCGAAAGTACACGTTTCATTGTGATTCAGAAAAAGACAAATAC 495
QY 2570 AGCATTTGCTTCAGAAATGGAAAAAGATAGGCAGTGGTCAATAAGATCTTACTCTGC 2629
DB 494 AGCATTTGCTTCAGAAATGGAAAAAGATAGGCAGTGGTCAATAAGATCTTACTCTGC 435
QY 2630 TCAAAACAGATGTCGCTGTTGCTGTTAAACAAATGGGAGGAGGATATCAAGAAAAAGCAA 2689
DB 434 TGAACACAGATGTCGCTGTTGCTGTTAAACAAATGGGAGGAGGATATCAAGAAAAAGCAA 375
QY 2690 AGAAATCGGAATTTGTGAACCCCATCATCCGGAAACAACTGTGGATAAACTTATCGTTTCGAA 2749
```



```

|||||
374 AGAACTGGAAATTGTGAACCCATCATCCGGAACAACTGTGTGAATAACTTATCGTTTGGAA 315
|||||
2750 ATACAAATTCGAATTTTCTTGGCATCTCATCATGTTGCTTCTTGGTACATCTCGTCCAGG 2809
|||||
314 ATACAAATTCGAATTTTCTTGGCATCTCATCATGTTGCTTCTTGGTACATCTCGTCCAGG 255
|||||
2810 ACATTACACATGTTATGATGACGATTAAGGATTAAGGATGAGCCAGATGAAGTCTATAAATGAC 2869
|||||
254 ACATTACACATGTTATGATGACGATTAAGGATTAAGGATGAGCCAGATGAAGTCTATAAATGAC 195
|||||
2870 CTACGGAATGCTTTTCTCTCTGCTAGATGTCGAAACCCCATCTCTGTTGCCCTGTTCCGGT 2929
|||||
194 CTACGGAATGCTTTTCTCTCTGCTAGATGTCGAAACCCCATCTCTGTTGCCCTGTTCCGGT 135
|||||
2930 TCATTATGCTCATTTATCATGTGAAAGCGAAAGAGCTTTATCGAACTTACAAAGGAACA 2989
|||||
134 TCATTATGCTCATTTATCATGTGAAAGCGAAAGAGCTTTATCGAACTTACAAAGGAACA 75
|||||
2990 TTACATCGGTGACTATGACAGCCAGGACTCGACAGAAATGGAACATTTTCTCCAAAC 3049
|||||
74 TTACATCGGTGACTATGACAGCCAGGACTCGACAGAAATGGAACATTTTCTCCAAAC 15
|||||
3050 TAACGTGAAGTACC 3063
|||||
14 TNACGTGAAGTACC 1

```

```

RESULT 13
BJ127233
LOCUS
DEFINITION
  BJ127233 unpublished oligo-capped cDNA library, C. elegans L1 stage
  Caenorhabditis elegans cDNA clone yk1344b04 5', mRNA sequence.
ACCESSION
  BJ127233
VERSION
  BJ127233.1 GI:18287390
KEYWORDS
  EST.
SOURCE
  Caenorhabditis elegans
  Caenorhabditis elegans
  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
  Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
  1 (bases 1 to 534)
AUTHORS
  Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
  and Sugano,S.
TITLE
  A complementary view of the C.elegans genome
JOURNAL
  Unpublished (2002)
COMMENT
  Contact: Tadasu Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
  Location/Qualifiers
    1..534
      /organism="Caenorhabditis elegans"
      /mol_type="mRNA"
      /strain="N2"
      /db_xref="taxon:6239"
      /clone="yk1344b04"
      /sex="hermaphrodite"
      /tissue_type="whole animal"
      /dev_stage="L1"
      /clone_lib="unpublished oligo-capped cDNA library, C.
      elegans L1 stage"

```

```

ORIGIN
Query Match          16.5%; Score 534; DB 2; Length 534;
Best Local Similarity 100.0%; Pred. NO. 2.8e-115;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1105 ATAAAGTCACCTGGTAAATCGGACAGAGGAAGAATGCAAAAAGTAGTACTACATTGT 1164
|||||
DB 1 ATAAAGTCACCTGGTAAATCGGACAGAGGAAGAATGCAAAAAGTAGTACTACATTGT 60

```

```

QY 1165 TCAAAATCTATGAGGAAAAAAGTTTCATGTAGTTTCCCACTACCACTAGTCAAG 1224
|||||
DB 61 TCAAAATCTATGAGGAAAAAAGTTTCATGTAGTTTCCCACTACCACTAGTCAAG 120
|||||
QY 1225 TTAAGAGTGAGCAAAAGATACGCTCTACCAATGGACATCTTGAAGTTTCATGAGAC 1284
|||||
DB 121 TTAAGAGTGAGCAAAAGATACGCTCTACCAATGGACATCTTGAAGTTTCATGAGAC 180
|||||
QY 1285 CACAAAGATACAAAGATCGAATTCGATCTGGTGATGCAAGCAAGTTTCTAAAGCGAGCTA 1344
|||||
DB 181 CACAAAGATACAAAGATCGAATTCGATCTGGTGATGCAAGCAAGTTTCTAAAGCGAGCTA 240
|||||
QY 1345 CACGAAACCTCAGCATACAAAGAAATACCCCTAAAGATGCTGAAAGAAATTTGATTTCT 1404
|||||
DB 241 CACGAAACCTCAGCATACAAAGAAATACCCCTAAAGATGCTGAAAGAAATTTGATTTCT 300
|||||
QY 1405 CTCTGGAAGAGCTAAATTTTGTGAAGATTTGGATTTATGCTCCAACTTCAGATGATCG 1464
|||||
DB 301 CTCTGGAAGAGCTAAATTTTGTGAAGATTTGGATTTATGCTCCAACTTCAGATGATCG 360
|||||
QY 1465 AATGTCAGGAAAGGTTTTCAGAGAGCAATGCTTGTGAATAGTGAATGAAACAATTA 1524
|||||
DB 361 AATGTCAGGAAAGGTTTTCAGAGAGCAATGCTTGTGAATAGTGAATGAAACAATTA 420
|||||
QY 1525 AATGACACAGTATTCGTGGATTTCAAGAAAAACAATTTGAATGTTGTTCCCGAAAAAG 1584
|||||
DB 421 AATGACACAGTATTCGTGGATTTTCAGAAAAACAATTTGAATGTTGTTCCCGAAAAAG 480
|||||
QY 1585 AACTTGTCTGCTGTTTGTGTAGTCAACGAAACAGCGGAAATCCATGCTTAG 1638
|||||
DB 481 AACTTGTCTGCTGTTTGTGTAGTCAACGAAACAGCGGAAATCCATGCTTAG 534

```

```

RESULT 14
BJ127087
LOCUS
DEFINITION
  BJ127087 unpublished oligo-capped cDNA library, C. elegans L1 stage
  Caenorhabditis elegans cDNA clone yk1342d11 5', mRNA sequence.
ACCESSION
  BJ127087
VERSION
  BJ127087.1 GI:18287244
KEYWORDS
  EST.
SOURCE
  Caenorhabditis elegans
  Caenorhabditis elegans
  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
  Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
  1 (bases 1 to 539)
AUTHORS
  Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
  and Sugano,S.
TITLE
  A complementary view of the C.elegans genome
JOURNAL
  Unpublished (2002)
COMMENT
  Contact: Tadasu Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
  Location/Qualifiers
    1..539
      /organism="Caenorhabditis elegans"
      /mol_type="mRNA"
      /strain="N2"
      /db_xref="taxon:6239"
      /clone="yk1342d11"
      /sex="hermaphrodite"
      /tissue_type="whole animal"
      /dev_stage="L1"
      /clone_lib="unpublished oligo-capped cDNA library, C.
      elegans L1 stage"

```

```

ORIGIN
Query Match          16.3%; Score 527; DB 2; Length 539;

```

Best Local Similarity 99.6%; Pred. No. 1.3e-113;  
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 935 AAAAATGATGGCGGAAAAATGACAAATCAGACAGCCGCGCCGCAAGAAATTCGACAAATT 994  
Db 1 AAAAATGATGGCGGAAAAATGACAAATCAGACAGCCGCGCCGCAAGAAATTCGACAAATT 60

QY 995 ATTGGAAAAATTTGAAGCTGAATGCGCAGAAATTTGGGATACGAAATGTCGAGATTGAC 1054  
Db 61 ATTGGAAAAATTTGAAGCTGAATGCGCAGAAATTTGGGATACGAAATGTCGAGATTGAC 120

QY 1055 AGAACGACATCTGACATTTCTAGATTGTGCGAGGAAAACTCTCTGTTTATAAAGTCTAC 1114  
Db 121 AGAACGACATCTGACATTTCTAGATTGTGCGAGGAAAACTCTCTGTTTATAAAGTCTAC 180

QY 1115 TGGTAAATCGGACAGAGGAGAAATGCAAAAAAGTACGATCTACATTTGTTCAAAATCTA 1174  
Db 181 TGGTAAATCGGACAGAGGAGAAATGCAAAAAAGTACGATCTACATTTGTTCAAAATCTA 240

QY 1175 TGAGGAAAACAAAAAGTTCAATTGAGTTTCCCACTACCACTAGTCAAAAGTTAAAAGTGG 1234  
Db 241 TGAGGAAAACAAAAAGTTCAATTGAGTTTCCCACTACCACTAGTCAAAAGTTAAAAGTGG 300

QY 1235 AGCAAAAGAAATCGCTGTACCAATGGAACATCTTGAAGTTTCATGAGAACGCCACAAAGATA 1294  
Db 301 AGCAAAAGAAATCGCTGTACCAATGGAACATCTTGAAGTTTCATGAGAACGCCACAAAGATA 360

QY 1295 CAAGAAATCGAATTTGATCTGGTGATGCAAGCAAGTTTCTAAAGCGAGCTACACGAAACC 1354  
Db 361 CAAGAAATCGAATTTGATCTGGTGATGCAAGCAAGTTTCTAAAGCGAGCTACACGAAACC 420

QY 1355 TCAGACTCAAAAGAAATACCCATAAAATGCTGAAAGAAATTTGGATTCTCTCTCTGAAGA 1414  
Db 421 TCAGACTCAAAAGAAATACCCATAAAATGCTGAAAGAAATTTGGATTCTCTCTCTGAAGA 480

QY 1415 GCTAAATTTGTTGAAAGAAATTTGGATTATGCTCCAACTTCAGATGATC 1463  
Db 481 GCTAAATTTGTTGAAAGAAATTTGGATTATGCTCCAACTTCAGATGATC 529

RESULT 15  
D27223/c  
LOCUS CELK002A7R Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA  
DEFINITION clone yk2a7 3', mRNA sequence.  
D27223  
D27223.1 GI:521294  
EST.  
Caenorhabditis elegans  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 449)  
REFERENCE Kohara,Y., Mitsuiki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and  
AUTHORS Tabara,H.  
TITLE Toward an expression map of the C.elegans genome  
JOURNAL Unpublished (1994)  
COMMENT Contact: Yuji Kohara  
Genome Biology Lab.  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.  
Location/Qualifiers  
1. .449  
/organism="Caenorhabditis elegans"  
/mol\_type="mRNA"  
/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone="yk2a7"  
/sex="hermaphrodite, male"  
/tissue\_type="whole animal"

/dev stage="varied"  
/clone\_lib="Yuji Kohara unpublished cDNA"

ORIGIN

Query Match 12.9%; Score 416.4; DB 9; Length 449;  
Best Local Similarity 99.0%; Pred. No. 2.1e-87;  
Matches 417; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2763 TTTTCTTGGGCAATCTCATCATGGTCTCTTGGTACATCTCGTCCAGGACATTACACTGTT 2822  
Db 421 TTTTCTTGGGCAATCTCATCATGGTCTCTTGGTACATCTCGTCCAGGACATTACACTGTT 362

QY 2823 ATGTATGACGATAAAGGAATGAGCCAGATGAAGTCTATAAAATGACCTACGACTTGT 2882  
Db 361 ATGTATGACGATAAAGGAATGAGCCAGATGAAGTCTATAAAATGACCTACGACTTGT 302

QY 2883 TTTTCTCTCTGTAGATGTGCAAAACCCATCTCGTTGCCCTGTTCCGGTTCAATTATGCTCAT 2942  
Db 301 TTTTCTCTCTGTAGATGTGCAAAACCCATCTCGTTGCCCTGTTCCGGTTCAATTATGCTCAT 242

QY 2943 TTATCATGTGAAAAAGCGAAAGAGCTTTATCGAACTTACAGGAACATTACATCGGTGAC 3002  
Db 241 TTATCATGTGAAAAAGCGAAAGAGCTTTATCGAACTTACAGGAACATTACATCGGTGAC 182

QY 3003 TATGCACAGCCACGAGCTCGACACGAAATGGAAACATTTTCTCCAAACTAAACGTGAAGTAC 3062  
Db 181 TATGCACAGCCACGAGCTCGACACGAAATGGAAACATTTTCTCCAAACTAAACGTGAAGTAC 122

QY 3063 CCTGGAATGTGTTCCGATTAACATTTTGCAAAAGTGTGCGCCGTTTCAATCAAAATTTTTC 3122  
Db 121 CCTGGAATGTGTTCCGATTAACATTTTGCAAAAGTGTGCGCCGTTTCAATCAAAATTTTTC 62

QY 3123 AATTGTAGATATTGTACTTACTTTTAAAGCCGGTTTCAAAATTTCAATCCATGAC 3182  
Db 61 AATTGTAGATATTGTACTTACTTTTAAAGCCGGTTTCAAAATTTCAATCCATGAC 2

QY 3183 T 3183  
Db 1 T 1

Search completed: July 6, 2006, 01:42:43  
Job time : 14574 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2006, 21:25:13 ; Search time 17487 Seconds  
(without alignments)  
11800.671 Million cell updates/sec

Title: US-10-645-746-2  
Perfect score: 3227  
Sequence: 1 cagcccaaaagtgtgaac.....tttaaaaaaaaaaaaaaa 3227

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_env.\*

2: gb\_pat.\*

3: gb\_ph.\*

4: gb\_pl.\*

5: gb\_pr.\*

6: gb\_ro.\*

7: gb\_sts.\*

8: gb\_sy.\*

9: gb\_un.\*

10: gb\_vi.\*

11: gb\_ov.\*

12: gb\_htg.\*

13: gb\_in.\*

14: gb\_om.\*

15: gb\_ba.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3227	100.0	3229	13	AF180730 Caenorhab
2	2602	80.6	39339	13	Z83113 Caenorhabdi
3	760.2	23.6	41648	13	AC084520 Caenorhab
4	129.6	4.0	30591	13	Z82085 Caenorhabdi
5	74.8	2.3	44042	4	AL023705 S.pombe c
6	69	2.1	4583	13	DQ178241 Caenorhab
7	69	2.1	29820	13	CE48F77 Caenorhabdi
8	68.8	2.1	25850	13	Z29121 Caenorhabdi
9	68.2	2.1	33706	13	AF016682 Caenorhab
10	64.8	2.0	4339	11	AJ719770 Gallus ga
11	63.4	2.0	3090	4	AK118258 Arabidops
12	62.4	1.9	734	2	BD214386 Novel hum
13	62.4	1.9	734	2	AK767980 Sequence
14	62.4	1.9	3040	5	BC025769 Homo sapi
15	62.4	1.9	3050	2	BD158050 Primer fo
16	62.4	1.9	3050	2	AX879833 Sequence
17	62.4	1.9	3050	5	AK022827 Homo sapi
18	62	1.9	7218	2	I66494 Sequence 14

19	61.8	1.9	1966	4	AK221864 Arabidops
20	61.4	1.9	771	2	AK551688 Sequence
21	61	1.9	1196	4	AK060907 Oryza sat
22	61	1.9	1196	4	AK104539 Oryza sat
23	61	1.9	1885	4	AK119248 Oryza sat
24	61	1.9	3178	4	AK065274 Oryza sat
25	61	1.9	3385	4	AK122078 Oryza sat
26	60.2	1.9	2897	4	AK122078 Oryza sat
27	59.2	1.8	38463	13	CEK01A6 Arabidops
28	58	1.8	3904	5	HS0803237 Caenorhabdi
29	57.6	1.8	2780	5	AB081473 Mus muscu
30	56.4	1.7	2144	5	AB168706 Macaca fa
31	56.4	1.7	3105	5	AK000040 Homo sapi
32	56.4	1.7	5990	5	AB046787 Homo sapi
33	55.4	1.7	1165	4	AB236789 Trifolium
34	54.8	1.7	2644	6	AB081474 Mus muscu
35	54.8	1.7	3080	6	AK129395 Mus muscu
36	54.8	1.7	3700	6	BC096023 Mus muscu
37	54.6	1.7	879	4	AY072819 Oryza sat
38	54.6	1.7	1211	4	AK101842 Oryza sat
39	54.6	1.7	1407	4	AY323483 Oryza sat
40	54.2	1.7	1141	2	AR579680 Sequence
41	54.2	1.7	1141	2	AX083744 Sequence
42	54.2	1.7	82312	12	AP007836 Lotus cor
43	53.8	1.7	4129	11	BC098982 Xenopus l
44	53.4	1.7	1522	13	AY701231 Orconecte
45	53	1.6	87618	4	AC007654 Genomic s

ALIGNMENTS

RESULT 1	AF180730	3229 bp	mRNA	linear	INV 07-NOV-1999
LOCUS	Caenorhabditis elegans RNA interference promoting factor RDE-1				
DEFINITION	(rde-1) mRNA, complete cds.				
ACCESSION	AF180730				
KEYWORDS	Caenorhabditis elegans				
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.				
REFERENCE	1 (bases 1 to 3229)				
AUTHORS	Tabata, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A. and Mello, C.C.				
TITLE	The rde-1 gene, RNA interference, and transposon silencing in C. elegans				
JOURNAL	Cell 99 (2), 123-132 (1999)				
PUBMED	10535731				
REFERENCE	2 (bases 1 to 3229)				
AUTHORS	Tabata, H., Sarkissian, M., Kelly, W.G., Grishok, A., Timmons, L., Fire, A. and Mello, C.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-AUG-1999) Medical School, Program in Molecular Medicine, University of Massachusetts, 373 Plantation Street, Worcester, MA 01605, USA				
FEATURES	Location/Qualifiers				
source	1..3229				
	/organism="Caenorhabditis elegans"				
	/mol_type="mRNA"				
	/db_xref="taxon:6239"				
	/chromosome="V"				
	/clone="yk296b10"				
gene	1..3229				
	/genes="rde-1"				
CDS	21..3083				
	/genes="rde-1"				
	/notes="similar to Caenorhabditis elegans K08H10.7"				
	/codon_start=1				
	/product="RNA interference promoting factor RDE-1"				
	/protein_id="AAF06159.1"				



Db 1801 |||||TCGGTAGATTGGAATAGCCGCAACAGNAGCGAAGATATGTTTGAAGCTCTCCCGGATA 1860  
Qy 1861 AAGAACAAAAGTCTTAATGTTTCATTATCAATTTCCAAACGACAACTGAATGCTTACGGTT 1920  
Db 1861 AAGAACAAAAGTCTTAATGTTTCATTATCAATTTCCAAACGACAACTGAATGCTTACGGTT 1920  
Qy 1921 TTGTGAACACATATTGCGATCACAACCATCGGTGAGCTTAATCAGCATATTAATTTCTGAAA 1980  
Db 1921 TTGTGAACACATATTGCGATCACAACCATCGGTGAGCTTAATCAGCATATTAATTTCTGAAA 1980  
Qy 1981 CAGTCACAAAAGCTTTGGCATCACTAAAGGCAACGAAAGGATCAAAACGAAATTTCTATC 2040  
Db 1981 CAGTCACAAAAGCTTTGGCATCACTAAAGGCAACGAAAGGATCAAAACGAAATTTCTATC 2040  
Qy 2041 AAATTCATTTGAAATCAACGCGAAATTAAGAGGTATTAACAGGAGCTTGACTGTCAG 2100  
Db 2041 AAATTCATTTGAAATCAACGCGAAATTAAGAGGTATTAACAGGAGCTTGACTGTCAG 2100  
Qy 2101 AAATTCAGAAATATCAACGAGAAAGAAAGAAAGACGGAACCAATGCCATTAACTATGT 2160  
Db 2101 AAATTCAGAAATATCAACGAGAAAGAAAGAAAGACGGAACCAATGCCATTAACTATGT 2160  
Qy 2161 ATGTTGGAATTTGATGTAACCTCATCAACCTCTACAGTGGAAATTTGATTTCTATAGCGG 2220  
Db 2161 ATGTTGGAATTTGATGTAACCTCATCAACCTCTACAGTGGAAATTTGATTTCTATAGCGG 2220  
Qy 2221 CTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGACTCAAG 2280  
Db 2221 CTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGACTCAAG 2280  
Qy 2281 AAGAAATGCTCCGGTGAGCGTGCAGTGCATGACGGGAAAGAAACAGATATTTTGG 2340  
Db 2281 AAGAAATGCTCCGGTGAGCGTGCAGTGCATGACGGGAAAGAAACAGATATTTTGG 2340  
Qy 2341 AAGCAAAAGTTTCGTGAAATTTGCTCAGAGAAATTCGACAGAAACAAACGACAAATCGAGCACAG 2400  
Db 2341 AAGCAAAAGTTTCGTGAAATTTGCTCAGAGAAATTCGACAGAAACAAACGACAAATCGAGCACAG 2400  
Qy 2401 CGCATATTTGATGTTCTATCGAGACGGAGTTAGCGATTCGAGAGTCTACGTGTTAGTCAATG 2460  
Db 2401 CGCATATTTGATGTTCTATCGAGACGGAGTTAGCGATTCGAGAGTCTACGTGTTAGTCAATG 2460  
Qy 2461 ATGAGCTTCGATCTTTAAAAGCGAAGTAACAAATTCATGTCGGAAACGGGATGGAGAAG 2520  
Db 2461 ATGAGCTTCGATCTTTAAAAGCGAAGTAACAAATTCATGTCGGAAACGGGATGGAGAAG 2520  
Qy 2521 ATCCAGAGCCGAAGTACACGTTCAATGTTGATTCAGAAACGACAAATACACGATTTGCTTC 2580  
Db 2521 ATCCAGAGCCGAAGTACACGTTCAATGTTGATTCAGAAACGACAAATACACGATTTGCTTC 2580  
Qy 2581 GAAGAAATGGAAAAGATAAGCCAGTGGTCAATAAAGATCTTACTCTCTGCTGAAACAGATG 2640  
Db 2581 GAAGAAATGGAAAAGATAAGCCAGTGGTCAATAAAGATCTTACTCTCTGCTGAAACAGATG 2640  
Qy 2641 TCCTGTTGCTGCTTTAAACAAATGGAGGAGGATGAAAGAAAGAAAGAAACAACTGGAA 2700  
Db 2641 TCCTGTTGCTGCTTTAAACAAATGGAGGAGGATGAAAGAAAGAAAGAAACAACTGGAA 2700  
Qy 2701 TTGTGAACCATCATCCGGAACCACTGTGATTAACCTTATCGTTTCGAAATACAAATTCG 2760  
Db 2701 TTGTGAACCATCATCCGGAACCACTGTGATTAACCTTATCGTTTCGAAATACAAATTCG 2760  
Qy 2761 ATTTTCTTTGGCATCTCATCATGTTGCTCTTGGTACATCTCGTCCAGGACATTAACACTG 2820  
Db 2761 ATTTTCTTTGGCATCTCATCATGTTGCTCTTGGTACATCTCGTCCAGGACATTAACACTG 2820  
Qy 2821 TTATGATGACGATAAAGGAATGAGCCAAAGATGAAGTCTATAAATGACCTACGGAATTG 2880  
Db 2821 TTATGATGACGATAAAGGAATGAGCCAAAGATGAAGTCTATAAATGACCTACGGAATTG 2880  
Qy 2881 CTTTCTCTCTGCTAGATGTCGAAAACCCATCTCGTTGCTGTTCCGGTTTCATTATGCTC 2940  
|||||

Db 2881 CTTTCTCTCTGCTAGATGTCGAAAACCCATCTCGTTGCCCTGTTCCGGTTTCATTATGCTC 2940  
Qy 2941 ATTTATCATGTGAAAAAGCGAAAGAGCTTTATCGAACTTACAAAGGAAACATTACATCGGTG 3000  
Db 2941 ATTTATCATGTGAAAAAGCGAAAGAGCTTTATCGAACTTACAAAGGAAACATTACATCGGTG 3000  
Qy 3001 ACTATGCACAGCCACGAGCTCGACACGAAATGGAACATTTTCTCAGAACTAACGTGAAGT 3060  
Db 3001 ACTATGCACAGCCACGAGCTCGACACGAAATGGAACATTTTCTCAGAACTAACGTGAAGT 3060  
Qy 3061 ACCCTGGAATGTCGTCGCATACATTTTGCAAAAGTGTGCCCGTTTCAATCAAAATTTT 3120  
Db 3061 ACCCTGGAATGTCGTCGCATACATTTTGCAAAAGTGTGCCCGTTTCAATCAAAATTTT 3120  
Qy 3121 TCAATTGTAGATATTTGACTTACTTTTAAAGCCCGGTTTCAAAAATTCATTCCCATG 3180  
Db 3121 TCAATTGTAGATATTTGACTTACTTTTAAAGCCCGGTTTCAAAAATTCATTCCCATG 3180  
Qy 3181 ACTAACGTTTTTCATAAATTTACTTGAATTTTAAAAAATTTAAAAAATTTAAAAA 3227  
Db 3181 ACTAACGTTTTTCATAAATTTACTTGAATTTTAAAAAATTTAAAAAATTTAAAAA 3227

RESULT 2  
CEK08H10/c

LOCUS CEK08H10 39339 bp DNA linear INV 09-AUG-2005  
DEFINITION Caenorhabditis elegans Cosmid K08H10, complete sequence.  
ACCESSION Z83113

VERSION Z83113.1 GI:3217648

KEYWORDS HTG.

SOURCE

ORGANISM

REFERENCE

AUTHORS

CONSTRM

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Jun 13, 1998 this sequence version replaced gi:1695070.  
Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.  
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.  
For a graphical representation of this sequence and its analysis see: <http://www.wormbase.org/perl/ace/elegans/seq/sequence?name=K08H10;class=Sequence>  
This sequence is the entire insert of clone K08H10. The true left end of clone K0688 is at 32990 in this sequence. The true right end of clone C2H6 is at 10790 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z81042. The end of this sequence (39233..39339) overlaps with the start of sequence Z73975.  
Location/Qualifiers  
1..39339 /organism="Caenorhabditis elegans"

FEATURES  
source

gene	<pre>/mol_type="genomic DNA" /strain="Bristol N2" /db_xref="taxon:6239" /chromosomes="V" /clone="K08H10" join(complement(638..744),complement(341..542), complement(105..294),complement(Z81042.1:24518..24775), complement(Z81042.1:24200..24465)) /locus_tag="C27H6.8" join(complement(638..744),complement(341..542), complement(105..294),complement(Z81042.1:24518..24775), complement(Z81042.1:24200..24465)) /locus_tag="C27H6.8" /standard_name="C27H6.8" /note="contains similarity to Pfam domain PF03690 (Uncharacterised protein family (UPF0160))" /codon_start=1 /product="Hypothetical protein C27H6.8" /protein_id="CAB05549.1" /db_xref="GI:3924825" /db_xref="InterPro:IPR003226" /translation="MTSLIGTHSGKPHCDFAFCFMLKOLPOFKDHSILRTRDAQLE KDIIVDVGSIIDHRSKQRYDHRGFTDMRTLEKINFDTKLSSAGLVVYHVGREVIN QILGNVSSMDLFLHRLYEQVESIDAIDNGISQYDGVPRYHSGNLSRSTQFNS HNWEPNDADERQQAMQFIIEFGRSVKILANWMPAREITIAAVDKRFDASGRI ILIEGNGCPKWHFFDIEVEKNADNNTIILFSDSTNASRWQAIPVDKMSFENRM PLFAAWRGURDDLSKESGIPGVFVHISGFTGGNLTREGAIAAMARKALEIGENPVK KAKIGN"</pre>	gene	<pre>/standard_name="K08H10.9" /note="contains similarity to Pfam domain PF04051 (transport protein particle (TRAPP) component, Bct3)" /codon_start=1 /product="Hypothetical protein K08H10.9" /protein_id="CAB05547.1" /db_xref="GI:3924823" /db_xref="InterPro:IPR007194" /db_xref="UniProtKB/TrEMBL:Q9XU81" /translation="MAFPFDILHLBIIRSTLESEKDRCAKEKFGTIINLNGPADIE KYESVQKAFLANAETKLESIGFRVGRQLVEKVSKEPKLVTELEIVFICKDFWSS VFGKVDNLRNTHQGVVYVQDGRFTLRSFPFGDQVVKESGYFLAPPCGLLRGALAGL NIRAIVTPTIESVPSVKFHIVQVQPK" complement(join(6724..6781,6827..6943,6995..7242, 7294..7435,7684..8083,8127..8425,8700..8875)) /locus_tag="K08H10.6" complement(join(6724..6781,6827..6943,6995..7242, 7294..7435,7684..8083,8127..8425,8700..8875)) /locus_tag="K08H10.6" /standard_name="K08H10.6" /note="contains similarity to Pfam domain PF00083 (Sugar and other) transporters)" /codon_start=1 /product="Hypothetical protein K08H10.6" /protein_id="CAB05545.1" /db_xref="GI:3924821" /db_xref="GOA:Q9XU83" /db_xref="InterPro:IPR005828" /db_xref="InterPro:IPR005829" /db_xref="InterPro:IPR007114" /db_xref="UniProtKB/TrEMBL:Q9XU83" /translation="NKVLPLENLTTPHLTTRLLATSVVATFAGGFHGYLISAVNP LSELIQQIIVNLRNYEYELSSLSLWSLACGLFAGMIGAYISIFLQSIQPR NSLLAAILQLFSPAPVFLAYSLRUCLELPIRSLISGLAFAGIGAPCVHLVNIITFH FRGLNLSGLIGNCAFLAATLTGTYLVGTQWKYFFIETPCVHLVNIITFH DSPTLLSIGIKESGSAIKVYGTGTCHIKKIMEDLRIDNGVKESLRQILKDKGCA QALSLAVINFSVAFGIVAFGTGTLQITGFSPEGSVANSLCSFASIVSALLAA TAMDKIGRPLLITSLVLALINI FMWSLVFLYDSTKDSFLAWPFLVLTFTVFSI GIGPAAVFIGELAPPQISIKMQSYSTSVQVQFAGSGICPIYLQLVESIGGFAMLFIV PLTITAAFYMYLPETKSGSPVEIYGLIQ" Join(17960..18067,18287..18427,18475..18790,19256..19589, 19640..19976,20048..20251) /locus_tag="K08H10.4" /standard_name="K08H10.4" /translation="MLFPAFSLILISFFLSLVVTTKTQYWCCHGDLVNNQHTCRPFT IVDAGSTGTRLHLVKFTHDPAIASHGMPFKVEKIFQVKPGLSFAKSPSSAADSL PELLARAKEVPHFWKEKPTITLKATAGRLPLGDMADDILSEVERIFNSGFFAAPP DAVNWPSDEGVSWFTNLILETLFTSHFGLNGLIARLNLQLETDNEIESTHQL PNEAVPSHEVGVERDIDFGHHIRLFTSHFGLNGLIARLNLQLETDNEIESTHQL ITSCMEPGQTEWYALKFNMINGSSSHSFESCYGTTKFNVESSIMHRELUKGSFV YLFSPFALNLSGLVKGNEGKIELRQFKAEEIACRRETEIDDGSHWMPQCIDL TVIYSLRDGYQFEDNQPLVLAKIKGMEVSWGQGLAFATANEFQLTEGAITALSSE PNSTVDQIFDLVSGTNOVLSYFNIIYV" Join(24915..24953,26257..26580) /locus_tag="K08H10.3" Join(24915..24953,26257..26580) /locus_tag="K08H10.3" /standard_name="K08H10.3"</pre>
CDS	<pre>complement(join(1035..1256,1333..1815,1877..2485, 2530..2644,2685..3296,3343..3559,3607..3785,3840..3895, 3940..4252,4295..4494,4550..4606)) /locus_tag="K08H10.7" complement(join(1035..1256,1333..1815,1877..2485, 2530..2644,2685..3296,3343..3559,3607..3785,3840..3895, 3940..4252,4295..4494,4550..4606)) /genes="rde-1" /standard_name="K08H10.7" /note="C. elegans RDE-1 protein; contains similarity to Pfam domains PF02170 (PAZ domain), PF02171 (Piwi domain)" /codon_start=1 /product="Hypothetical protein K08H10.7" /protein_id="CAB05546.2" /db_xref="GI:14530521" /db_xref="InterPro:IPR003165" /db_xref="InterPro:IPR003165" /db_xref="UniProtKB/TrEMBL:Q9XU82" /translation="MSSNFPLEKGFYRSLDPEMKWLARPTGCKDGKYEKKVLLLV NPKFSKKIDREYIEYKMTKEVNLKAPGFPFKTEIPIDRKLFWQHLRHEKK QTDFILEDVFEKDTYVSVCLNLTVTSKLVSEKVVKKDSEKDEKLEKILYTI LTYRKPHLNFSENPKEDEANRYKFLKNYMTQKRYAPFVNEEIKVQFARNFYD NNSILKVPFDPNRFQESLEVAIPRIEAWFGYIGIKELFDGEPNFAIVDKLFYN APWMSLDVLLITVDQSCNDDVRKDKTKLMAGKWTIRQAARPIROLLENLKLCA EVDNEMSLRTEHLYFLDLCENSLVYKVTGSKDRGNKAKYDTTLFKIYENKKFI EPFHLPLVKSGAEYVMEHLEVHEKPIKRIKIDLVMDQKFLKRAIRKPHDKKE NTLKMELDFSEELNLFVERFGLSKLQMIIECPKVLKPELNVSNVQIKRTFVIR GFOEKQNLVNPBKELCACVAVNTEAGNCLNEENVVKFYTELIGCKFRGIRIGANE NRGAQNLVYDQATNEVAYKNTLTNTIGRFAIAEAKNMPERLPDKEQKLMFTII SKROLNAGYKVKHCDHTIGVANQHTTSFTVTKALASRHEGSKRIFYQIALKNVIA LGINQELDWSIAEISLPEKERRKTMPLTNTVGDIVTPTSYSGIDYIAAVASIN PGGTIYRNMIIVTOEBCRGERAVAHGRENTDILEAKFKLLREFANNRNRAPIHVV YRIGVDSMLRVSHDELRLSLKSEVKQFMSERDGEDPEPKYTFIVQKRNHTRLLRM EKDKPVNKDLTPAETDVAVAKWEEDMKESKETGIWNPSSGTTVDKLIYVKYKFD FLASHGVLTGRSPGHTYMTDDKMSQDEVYKMTYGLAFLASRCRKEPISLPVPVHY AHLSCFKAELRYTKYEHYIGDYAOPRTREMEHFLOTNVKYVPGMSFA" Join(4899..4994,5051..5122,5167..5268,5318..5395, 5444..5584,5630..5698) /locus_tag="K08H10.9" Join(4899..4994,5051..5122,5167..5268,5318..5395, 5444..5584,5630..5698) /locus_tag="K08H10.9"</pre>	CDS	<pre>/standard_name="K08H10.9" /note="contains similarity to Pfam domain PF01150 (GDA)/CD39 (nucleoside phosphatase) family)" /codon_start=1 /product="Hypothetical protein K08H10.4" /protein_id="CAB05544.1" /db_xref="GI:3924820" /db_xref="GOA:Q9XU84" /db_xref="InterPro:IPR000407" /db_xref="UniProtKB/TrEMBL:Q9XU84" /translation="MLFPAFSLILISFFLSLVVTTKTQYWCCHGDLVNNQHTCRPFT IVDAGSTGTRLHLVKFTHDPAIASHGMPFKVEKIFQVKPGLSFAKSPSSAADSL PELLARAKEVPHFWKEKPTITLKATAGRLPLGDMADDILSEVERIFNSGFFAAPP DAVNWPSDEGVSWFTNLILETLFTSHFGLNGLIARLNLQLETDNEIESTHQL PNEAVPSHEVGVERDIDFGHHIRLFTSHFGLNGLIARLNLQLETDNEIESTHQL ITSCMEPGQTEWYALKFNMINGSSSHSFESCYGTTKFNVESSIMHRELUKGSFV YLFSPFALNLSGLVKGNEGKIELRQFKAEEIACRRETEIDDGSHWMPQCIDL TVIYSLRDGYQFEDNQPLVLAKIKGMEVSWGQGLAFATANEFQLTEGAITALSSE PNSTVDQIFDLVSGTNOVLSYFNIIYV" Join(24915..24953,26257..26580) /locus_tag="K08H10.3" Join(24915..24953,26257..26580) /locus_tag="K08H10.3" /standard_name="K08H10.3"</pre>



```
/note="contains similarity to Pteronotus parnellii Brcal
(Fragment).; TR:Q7YQ1"
/codon_start=1
/product="Hypothetical protein K08H10.3"
/protein_id="CAB05542.1"
/db_xref="GI:3924818"

Query Match      80.6%; Score 2602; DB 13; Length 39339;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 3211; Conservative 0; Mismatches 0; Indels 509; Gaps 10;

QY      1  CAGCCACAAAGTGATGAAACATGCTCTCGAATTTTCCGGAATTTCCGGAATTTGGAAGAGGATTTTATC 60
DB      4626 CAGCCACAAAGTGATGAAACATGCTCTCGAATTTTCCGGAATTTTCCGGAATTTGGAAGAGGATTTTATC 4567

QY      61  GTCATTCTCTCGATCC-----76
DB      4566 GTCATTCTCTCGATCCGGTATGATCAATTTATAGCAGCTATAGATATATAAGTTTGATA 4507

QY      77  -----GGAGATGAAATGGCTTCGAGAGCCCACTGGTAAATCGGACGGCAAAATTC 125
DB      4506 TTAATATTAATAGGAGATGAAATGGCTTCGAGAGCCCACTGGTAAATCGGACGGCAAAATTC 4447

QY      126 TATGAGAGAAAGTACTCTTTTGTAAATTTGGTAAATTTGGTCAAGTTCTCCAGCAAAATTTACGAT 185
DB      4446 TATGAGAGAAAGTACTCTTTTGTAAATTTGGTAAATTTGGTCAAGTTCTCCAGCAAAATTTACGAT 4387

QY      186 CGGGAATACTACGAGTATGAAGTGAAATGACAAAGGAAGTATTGAAATAGAAACACGAGGA 245
DB      4386 CGGGAATACTACGAGTATGAAGTGAAATGACAAAGGAAGTATTGAAATAGAAACACGAGGA 4327

QY      246 AAACCTTTCCAAAAAAGACAGAAATTTCCAAAT-----277
DB      4326 AAACCTTTCCAAAAAAGACAGAAATTTCCAAATTTGAATGATGCTGTGTAATTTAGTCAAAACTA 4267

QY      278 -----TCCGATCGTGCAGAAACTCTCTGCGCAACATCTTCGGCATGAGAAG 323
DB      4266 ATTTTATTTTTCAGTCCCGATGTCGCAAACTCTCTGCGCAACATCTTCGGCATGAGAAG 4207

QY      324 AAGCAGACAGATTTTATTTCTCGAAGACTATGTTTTGATGAAAGGACACTGTTTATAGT 383
DB      4206 AAGCAGACAGATTTTATTTCTCGAAGACTATGTTTTGATGAAAGGACACTGTTTATAGT 4147

QY      384 GTTTGTGCACTGAACACTGTGCATCAAAATGCTGTTTCGAGAGAAAGTAGTAAAG 443
DB      4146 GTTTGTGCACTGAACACTGTGCATCAAAATGCTGTTTCGAGAGAAAGTAGTAAAG 4087

QY      444 GATTCGGAGAAAAGATGAAGAGGATTTGAGAAAAGATTTCTATACAAATGATACTT 503
DB      4086 GATTCGGAGAAAAGATGAAGAGGATTTGAGAAAAGATTTCTATACAAATGATACTT 4027

QY      504 ACCTATCGTAAAAATTTTCACTGAACTTTAGTCGAGAAAATCGGAAAAAGACGAGAA 563
DB      4026 ACCTATCGTAAAAATTTTCACTGAACTTTAGTCGAGAAAATCGGAAAAAGACGAGAA 3967

QY      564 GCGAATCGGAGTTACAAATTCCTGA-----588
DB      3966 GCGAATCGGAGTTACAAATTCCTGAAGTTTATGAAAAACACGCAATTTATAACAAACAAA 3907

QY      589 -----AGAAATGTTATGACCCGAGAAAGTTTCGCTACGCGCTTTTGTGAAAGGAGGA 639
DB      3906 TTAGCTTTGAGAATGTTATGACCCGAGAAAGTTTCGCTACGCGCTTTTGTGAAAGGAGGA 3847

QY      640 TTAAGT-----646
DB      3846 TTAAGTGTGAGTTGCAATTAATAATAATACCTCACTCACTTATATATTTAA 3787

QY      647 -ACAAATCGGAAAAATTTTGTGTGACGATTAATTAATTTCTCGAGTTCTCGAATCGT 705
DB      3786 GACAAATTCGAAAAATTTTGTGTGACGATTAATAATTTCTCGAGTTCTCGAATCGT 3727

QY      706 TTCACGATCCAAACAGATTCGAACATCATTAGAGTAGCAACCAAGATCGAAGCATGGT 765
DB      706 AGTAAAGTTCTACACCGAACTTAATTTGGTGGTTGCAAGTTCCGTTGGATACCGAATTTGGTGCC 2587
```

```
DB      3726 TTCAGATCCAAACAGATTCGAACATCATTTAGAAGTAGCACCAAGAAATCGAAGCATGGT 3667
QY      766 TTGGAATTTACATTTGGAATCAAGAAATTTGTTTCGATGGTGAACCTGTGCTCAATTTTGCAA 825
DB      3666 TTGGAATTTACATTTGGAATCAAGAAATTTGTTTCGATGGTGAACCTGTGCTCAATTTTGCAA 3607
QY      826 -----TTGTCGATAAACT 838
DB      3606 GTAAGTTTGAAAACTGCGATAAAAAATCATGTGATTTTGTGAAGTTGTGATAAACT 3547
QY      839 ATTCTACAATGACCGAAAAATGCTCTCTCTGGATATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 898
DB      3546 ATTCTACAATGACCGAAAAATGCTCTCTCTGGATATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3487
QY      899 GTCGTGTAAACGATGATGACGAAAAAGATCTTAAAAACAAACTGATGCGGGAAAAAATGAC 958
DB      3486 GTCGTGTAAACGATGATGACGAAAAAGATCTTAAAAACAAACTGATGCGGGAAAAAATGAC 3427
QY      959 AATCAGACAAGCCGCGCGGCAAGAAATTCACAAATTTTGGAAAAATTTGAAGCTGAAATG 1018
DB      3426 AATCAGACAAGCCGCGCGGCAAGAAATTCACAAATTTTGAAGCTGAAATG 3367
QY      1019 CGCAGAGTTTGGGATAACCGAAT-----1042
DB      3366 CGCAGAGTTTGGGATAACCGAATTTAGTTTAAATTTATTCAAACAATTTATATACAAAT 3307
QY      1043 -----GTGAGATTGACAGAACGACATCTGACATTTCTAGATTTTGTGCGAGGAAA 1092
DB      3306 TGATTTTCAGTTCGAGATTCACAGAACGACATCTGACATTTCTAGATTTTGTGCGAGGAAA 3247
QY      1093 ACTCTCTGTTTATAAGTCACTGGTAAATTCGACAGAGGAAATTCGAAAAAAGTAGC 1152
DB      3246 ACTCTCTGTTTATAAGTCACTGGTAAATTCGACAGAGGAAATTCGAAAAAAGTAGC 3187
QY      1153 ATACTACATTTTCAAAATCTATGAGAAAACAAAAGTTTCATTTAGTTTCCCACTAC 1212
DB      3186 ATACTACATTTTCAAAATCTATGAGAAAACAAAAGTTTCATTTAGTTTCCCACTAC 3127
QY      1213 CACTAGTCAAAGTTAAAAGTGGAGCAAAAGAAATACGCTGTACCAATGGAAATCTTGAAG 1272
DB      3126 CACTAGTCAAAGTTAAAAGTGGAGCAAAAGAAATACGCTGTACCAATGGAAATCTTGAAG 3067
QY      1273 TTCAAGAGAGCCACAAAGATAACAAGATTCGAATTTGATCTGGTGAATGCAAGACAAAGTTTC 1332
DB      3066 TTCAAGAGAGCCACAAAGATAACAAGATTCGAATTTGATCTGGTGAATGCAAGACAAAGTTTC 3007
QY      1333 TAAAGCGAGCTACACGAAACCTCAGACCTACAAAGAAATACCTTAAATGCTGGAAG 1392
DB      3006 TAAAGCGAGCTACACGAAACCTCAGACCTACAAAGAAATACCTTAAATGCTGGAAG 2947
QY      1393 AATTGGATTTCTCTCTGAGAGCTAAAATTTTCTGTAAGATTTTGAAGATTTGATTTCTCCAAAC 1452
DB      2946 AATTGGATTTCTCTCTGAGAGCTAAAATTTTCTGTAAGATTTGATTTCTCCAAAC 2887
QY      1453 TTCAAGATGATGCAATGTTCAGGAAAGGTTTGAAGAGCAATGCTTTGTGAATAGTGAA 1512
DB      2886 TTCAAGATGATGCAATGTTCAGGAAAGGTTTGAAGAGCAATGCTTTGTGAATAGTGAA 2827
QY      1513 ATGAACAAATTTAAATGACACCAAGTTCGTCGATTTTCAAGAAAACAAATTTGAATGCG 1572
DB      2826 ATGAACAAATTTAAATGACACCAAGTTCGTCGATTTTCAAGAAAACAAATTTGAATGCG 2767
QY      1573 TTCCGAAAAAAGAACTTTTGTGCTGCTTTTCTAGTCAACGAAACAGCGGGAATCCAT 1632
DB      2766 TTCCGAAAAAAGAACTTTTGTGCTGCTTTTCTAGTCAACGAAACAGCGGGAATCCAT 2707
QY      1633 GCTTAGAAGAGAACGAGTTGT-----1654
DB      2706 GCTTAGAAGAGAACGAGTTGTGTAAAGTTTCTACGTAGATTTATTCGAAATATTTTC 2647
QY      1655 --TAAAGTTCTACACCGAACTTAATTTGGTGGTTGCAAGTTCCGTTGGATACGAAATTTGGTGC 1712
DB      2646 AGTAAAGTTCTACACCGAACTTAATTTGGTGGTTGCAAGTTCCGTTGGATACGAAATTTGGTGC 2587
```





```

source
1. .41648
/organism="Caenorhabditis briggsae"
/mol_type="genomic DNA"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="G18K16"

ORIGIN
Query Match      23.6%; Score 760.2; DB 13; Length 41648;
Best Local Similarity 59.5%; Pred. No. 3.3e-208;
Matches 1596; Conservative 0; Mismatches 803; Indels 284; Gaps 7;

Qy 648 CAATTCGGAAAAATTTTGTGTGACGATAATAATTCAAATTCCTGCGAGTTCCTGAATCGTTT 707
Db 9502 CAACTTCGCAAAAATTTTCATCTACGATGGAACTCCCAATTCCTCGAGTTCAGAAATCTTTT 9561
Qy 708 CACGATCCAAACAGATTCGACCAATCATTTAGAAGTAGCACCAAGATCCGAAGCATGGTTT 767
Db 9562 CACGACCCTGATCGATTTTGAACATCTCTGGAATCGCGCCAGCATGAATCTTGGTTC 9622
Qy 768 GGAATTTTACATTCGAATCAAAGAATTTGTCGATGGTGAACCTGTGTCTCAATTTTG --- 822
Db 9622 GGAATCTATATCGCTGTTAAAGAACTTTTCGACGGAGACCCGGTTCCTCAATTTGCCAGT 9681
Qy 823 -----CAATTTGTCTGATAAACT 838
Db 9682 GAGAAATTAATAATTTTAAAGCTGAAAAATATTAACAATTAATTTTTCAGTTATTGACAAAT 9741
Qy 839 ATTCTCAATGCAACCGAATATGTCCTTTCTGGATATCTTCTCTAATTTGTGCAGCCCCA 898
Db 9742 GTTCTCAACGCTCCCAAAATGTCCTCTCTCGACTACCTGCTTCTGATTTGTGGATCCGCA 9801
Qy 899 GTCGTGTAAACGATGATGTACGAAAAGATCTTAAACAAAACTGATGCGCGGAAAAATGAC 958
Db 9802 ATCTCATATGCAAGAGTTTCAAGAAACCAATAAAGAACAGTTGAAATCGGHAATTTGAC 9861
Qy 959 AATCAGACAAGCGCGCGGCAAGAAATTCGACAAATTTATGGAAAAATTTGAAGCTGAAATG 1018
Db 9862 AATTAAGCCAGCTATGTTTCAAAGCATGATCAATTTGCTTGGCAATTTTAAAGCTTAAAGC 9921
Qy 1019 CGCAGAAGTTTGGGTAACGAATATGTCGAA ----- 1049
Db 9922 AGCATCGGTATGGAATCTTAAAGAGGTATGAACTACTATTGAAGAAATGAGACGTTAATCA 9981
Qy 1050 -----TTGACAGAAACGACATCTGACATTTCTAGATTTCTGTGCGAGGA 1090
Db 9982 AAATATTTCAGTCCAGAGTTTCTGAAAGTCATCTGACATTTATCAGATTCAGACAGTA 10041
Qy 1091 AAATCTCTTGTTTTAAAGTCACTGGTAAATCGGACAGAGGAAGAAATGCAAAAAAGTA 1150
Db 10042 TAATTCACGAAGAATGATTCACGTTTCCAAGAGGAAGACAGACAGAAATGCCCAATTTGA 10101
Qy 1151 CGATCTACATTTGTTTCAAATCTATGAGGAAACAAAAAGTTTCATTTGAGTTTCCCACCT 1210
Db 10102 CAGAGTTCCTATGTTTCAAATTTATGAGAAGAACCGTAAATACATCGAGTTTCTCTCGCTT 10161
Qy 1211 ACCACTAGTCAAAGTTTAAAGTGGACCAAGAATACGCTGTACCAATCGGAACATCTTTGA 1270
Db 10162 ACCATTGGTGGTGTGAAGCGGTAATAACGAGTATTCGGTCCCAATGGAGTTTCTTAGA 10221
Qy 1271 AGTTTCATGAGAAGCCCAAGAGATACAAGAAATCGAATTTGATCTGGTGTATGCAAGACAAGTT 1330
Db 10222 AGTGTCGAAAAACCAACGCTACAAAATCGAATCGACTTTTGGATGCAAGACAAAT 10281
Qy 1331 TCTTAAGCGAGCTACACGAAAACCTCAGCTACAAAGAAATAACCTTAAATAATGCTGAA 1390
Db 10282 TGTGAAATGACGACACGCGGATCTCTCATATGATATAAGAGGAAGCGCTTGAGATGTTGAA 10341
Qy 1391 AGAATTTGGAATTTCTCTCTGMAAGAGCTAAATTTGTTGGAAGATTTGCAATATGCTCAA 1450
Db 10342 AGATCTCGAATTTTTCATCTGGATCGCTGAACTTTGTTGAACGGTTTGAAATCAGTACCGA 10401
Qy 1451 ACTTCAGATGATCGAATGTCGAGGAAGGTTTTTGAAGAGGACCAATGCTTTGTGAAATAGTGT 1510

```

```
Db 11482 ACGGTGTGACGACTCAGAAATGCTCGTGTGCTGCTCATGATGATTAAGGTCAATTTGAAG 11541
Qy 2482 GCGAAGTAAACAAATTCATGTCGAAACGGGATGGAGAAGATCCAGAGCCGGAAGTACACCT 2541
Db 11542 CGGAACTTGAACGATTCCTGAGGAGCGAGGCAAGAAATGAGCAAAATTCGGAGTACACCT 11601
Qy 2542 TCATTTGTGATTCGAAAACACACATACAGGATTCGTCGAGATGCGAAGATGGAAGAAAGATAAGC 2601
Db 11602 TCATCGTTCTACAAAAGACACAAAACTCGGCTTTTCGCAAGATTTGAAATGAGCGCG 11661
Qy 2602 CAGTGTCAATAAAGATCTTACTCTGCTGMAACAGATGTCGCTGTGCTGTGCTGTTAAAC 2661
Db 11662 C-----GAAAGATGAGGAGCAGCGGAAAC 11685
Qy 2662 AATGGAGGAGGATATGAAGAAAGCAAACTCGAATTTGGAACCCATCATCCGGAA 2721
Db 11686 GATGGACGAAGAGCTGAAGGCTAGCCAAAACACGGGTTTGTGAATCTTACTTCTGGGA 11745
Qy 2722 CAACTGTGNTAAACTATGCTTTTGGAAATACAAATTCGATTTTCTTGGCATCTCATC 2781
Db 11746 CAACTGTAGTAACTATAGTTTCCAAATATAGTTTCCAGTATAGTTTCCGCTTCTCTCATC 11805
Qy 2782 ATGCTGCTCTGTGATCATCTCGTCCAGGACATTACACTGTTATGTATGACGATAAAGGAA 2841
Db 11806 ACGTTCGGCTTGGAAATCGGACCTGGTTATTACACAGTTATGATGACGACGACGAA 11865
Qy 2842 TGAGCCCAAGATGAAGTCTAT----- 2861
Db 11866 TGACAAAGGATGAATATATGTGAGTTCCGAGAGAAAAATAGTTTCCACGAACTATTTTC 11925
Qy 2862 --AAATGACCTACGACTGCTTTCTCTGCTAGATGTCGAAACCCATCTCGTTCG 2919
Db 11926 AGAAATGACCTACGAAATGTCTATCTTCGCTAGATGCGGTAAACCAATTTCTCTAC 11985
Qy 2920 CTGTCCTCGTTCAATATGCTCATTTATCATGTGAAAAAGCGAAAGACTTTATCGAACTT 2979
Db 11986 CAACACCGTTCTACGCTCATCTCTTGTGAGAGGCGAAGAGCTATATCGCTGT 12045
Qy 2980 ACAAGGAACATTACATCGGTGACTATGCAAGCCAGGACTCGACAGAAATGGAACATT 3039
Db 12046 TCAAGAATACGGTTTGCACACTTTCCACGACAGGAGGAGGAGATATCGAACAT 12105
Qy 3040 TTCTCCAACTACGTTGAGTACCTCGGATGCTGCTTCCATA 3082
Db 12106 ACCTGCAAAACCAATATCCAGTATCCTCGGATGCTTTCGCTA 12148

RESULT 4
CEZK218
LOCUS CEZK218 30591 bp DNA linear INV 25-AUG-2005
DEFINITION Caenorhabditis elegans Cosmid ZK218, complete sequence.
ACCESSION Z82085
VERSION Z82085.1 GI:1753056
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
          Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
          1 (bases 1 to 30591)
REFERENCE
  1. C. elegans Sequencing Consortium
  2. Genome sequence of the nematode C. elegans: a platform for
  3. investigating biology
  4. Science 282 (5396), 2012-2018 (1998)
  5. PUBLISHED 9851916
  6. 2 (bases 1 to 30591)
  7. McMurray, A.A.
  8. Direct Submission
  9. Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger
  10. Institute, Hinxton, Cambridge CB10 1SA, England and Department of
  11. Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
  12. worm@sanger.ac.uk
  13. On Dec 25, 1996 this sequence version replaced gi:1695121.

COMMENT
```

Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

For a graphical representation of this sequence and its analysis see: <http://www.wormbase.org/perl/ace/elegans/seq/sequence?name=ZK218;class=Sequence>

This sequence is the entire insert of clone ZK218. The true right end of clone T05B12 is at 10829 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z81585.

The end of this sequence (30488..30591) overlaps with the start of sequence AL031628.

Location/Qualifiers

1..30591

/organism="Caenorhabditis elegans"

/mol\_type="genomic DNA"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="V"

/clone="ZK218"

/join(1174..1278,1321..1585,1638..2005)

/locus\_tag="ZK218.1"

/join(1174..1278,1321..1585,1638..2005)

/locus\_tag="ZK218.1"

/standard\_name="ZK218.1"

/note="contains similarity to Pfam domain PF01549 (SHTK domain)"

/codon\_start=1

/product="Hypothetical protein ZK218.1"

/protein\_id="CAB04984.1"

/db\_xref="GI:3881559"

/db\_xref="InterPro:IPR003582"

/translation="MIAPFVILTVLAFVQAVQVPERNCTAYDAVATDFVYTPGATN CVNTYSDKTSLSYIAAANAPVPLGLKTLRPGKFTTAATAVTSPIDADLVRAAVS TCPKTCGYCCSPDYNCPVQFPLNCATILPSOCTNQOFRVIAQDCPSACGLCNAG GCVDAIMGCANDISICNVGMQDFVNLNCQRTCSRSSAPVPGGCTYNRDSGTFCAAW AANGFCQNEFSPDRKQYCATCKIC"

/complement(join(6531..6898,6950..7208,7254..7370))

/locus\_tag="ZK218.3"

/complement(join(6531..6898,6950..7208,7254..7370))

/locus\_tag="ZK218.3"

/standard\_name="ZK218.3"

/note="contains similarity to Pfam domain PF01549 (SHTK domain)"

/codon\_start=1

/product="Hypothetical protein ZK218.3"

/protein\_id="CAB04982.1"

/db\_xref="GI:3881557"

/db\_xref="InterPro:IPR003582"

/translation="WVASIVLVVLAFCVQTPQPAIGSELNCTVYSTTVPDVFVYTP GATNCVNAYADSTCASLYAITTDEPAFVPGTSPRLKCTGAGEAIGADPDLVRAA VSTCPKTCGFCQCSADYNPVQFPLNCATILPSOCTNQOFRVIAQDCPSACGLCNAG GCGCVDAVMGCANDISICNVGMQDFVNLNCQRTCSRSSSTTVTGCTYNRDSSTACA AWAIVGFCQNDPFTPAQRKQYCATCKIC"

/join(8325..8714,8764..8811,8864..8968)

/locus\_tag="ZK218.4"

/join(8325..8714,8764..8811,8864..8968)

/locus\_tag="ZK218.4"

/standard\_name="ZK218.4"

/note="contains similarity to Reclinomonas americana Subunit of ABC transporter for cytochrome c1.; TR:O21278"

/codon\_start=1

/product="Hypothetical protein ZK218.4"

/protein\_id="CAB04985.1"

/db\_xref="GI:3881560"



Mungall,K., Murphv,L., Niblett,D., Odell,C., Oliver,K., O'Neil,S., Pearson,D., Quail,M.A., Rabbintowitch,E., Rutherford,K., Rutter,S., Saunders,D., Seeger,K., Sharp,S., Skelton,J., Simmonds,M., Squares,R., Squares,S., Stevens,K., Taylor,K., Taylor,R.G., Tivey,A., Walsh,S., Warren,T., Whitehead,S., Woodward,J., Volckaert,G., Aert,R., Robben,J., Grymonprez,B., Meltjens,I., Vanstreels,E., Rieger,M., Schafer,M., Muller-Auer,S., Gabel,C., Fuchs,M., Dusterhoft,A., Fritzc,C., Holzer,E., Moestl,D., Hilbert,H., Borzým,K., Langer,I., Beck,A., Lehrach,H., Reinhardt,R., Pohl,T.M., Eger,P., Zimmermann,W., Wedler,H., Wambutt,R., Purnelle,B., Goffeau,A., Cadieu,B., Dreano,S., Gloux,S., Lelaure,V., Mottier,S., Galibert,F., Aves,S.J., Xiang,Z., Hunt,C., Moore,K., Hurst,S.M., Lucas,M., Rochet,M., Gaillardin,C., Tallada,V.A., Garçon,A., Thode,G., Daga,R.R., Cruzado,L., Jimenez,J., Sanchez,M., del Rey,F., Benito,J., Dominguez,A., Revuelta,J.L., Moreno,S., Armstrong,J., Foréburg,S.L., Cerutti,L., Lowe,T., McCombie,W.R., Paulsen,I., Potashkin,J., Shpakovski,G.V., Ussery,D., Barrell,B.G. and Nurse,P.

The genome sequence of Schizosaccharomyces pombe  
Nature 415 (6874), 871-880 (2002)

11859360

2 (bases 1 to 44042)

Wood,V., Rajandream,M.A., Barrell,B.G., Murphy,L. and Harris,D.

Direct Submission

Submitted (27-MAY-1998) European Schizosaccharomyces genome  
sequencing project, Sanger Institute, The Wellcome Trust Genome  
Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk

Notes:

Details of S. pombe sequencing at the Sanger Institute are

available on the World Wide Web.

(URL, [http://www.sanger.ac.uk/Projects/S\\_pombe/](http://www.sanger.ac.uk/Projects/S_pombe/)) (URL,

<http://www.sanger.ac.uk/genedb/pombe/index.jsp>) (URL,

CDS are numbered using the following system eg SPAC5H10.01c. SP (S.  
pombe), A (chromosome 1), cSH10 (cosmid name), .01 (first CDS), c  
(complementary strand). However, clones may have been reorientated  
since the original submission, therefore the complementary strand  
notation may be invalid for strand inference. IMPORTANT: This  
sequence MAY NOT be the entire insert of the sequenced clone. It  
may be shorter because we only sequence overlapping sections once.

Location/Qualifiers

#### FEATURES

source

1. .44042

/organism="Schizosaccharomyces pombe"

/mol\_type="genomic DNA"

/strain="972h"

/db\_xref="taxon:4896"

/chromosome="III"

/map="III"

/clone="cosmid c736"

<1. .1158

/gene="SPCC553.01c"

/note="synonym: SPCC736.01c"

<1. .1158

/gene="SPCC553.01c"

/codon\_start=1

/protein\_id="CAAL19265.1"

/db\_xref="UniProtKB/Swiss-Prot:O74939"

/translation="DHQDQSSVAEPTSKRLRLTLPLILRKPTGRQNFIVT  
TPSKNLYRSVKTAELASLQNTTETDNEIPNSPDFLDAASEDEHIPSLSLSLL  
IFLQKNKAFQVPRVPSFSPPLVKPASSSFGSRHSHLGTSLRNSLNGSSASS  
TFLQNNVNPVSPSTVLEHAKHPSLAKHSGVSLRTGLLIYPKYANGSKRKGFGAT  
SIISGEHSLPKNSKRVSVSPFNVPSTVSSDQVQKKGKGNVPAIETDSSPS  
DTISSPTVRANVLSKQKPKILTKIPDIFSKAGKSPILPTFQESCLPIEPFS  
SNITDLPFVNTVLASTSSKIKRAKLMLKFKRAMTQTQGRDLSDSDGDFI"  
complement(1. .181)  
/notes="nominal overlap with cosmid SPCC553, EM:AL023704 S.  
pombe chromosome 3"

misc\_feature

1655. .2515

/gene="SPCC736.02"

1655. .2515

/gene="SPCC736.02"

/notes="hypothetical protein; sequence orphan"

/codon\_start=1

gene  
complement(2836. .4125)  
/gene="SPCC736.03c"  
CDS  
complement(2836. .4125)  
/gene="SPCC736.03c"  
/note="phenylalanyl-tRNA synthetase; localization  
mitochondrion (predicted); similar to S. cerevisiae MSF1"  
/codon\_start=1  
/protein\_id="CAAL19267.1"  
/db\_xref="GI:3169073"  
/db\_xref="GOA:O74952"  
/db\_xref="InterPro:IPR002319"  
/db\_xref="InterPro:IPR004530"  
/db\_xref="InterPro:IPR005121"  
/db\_xref="InterPro:IPR006195"  
/db\_xref="UniProtKB/TrEMBL:O74952"  
/translation="MPPHRVSHLSKLFRRFPNWKRRKVQTDSSNVPFIHSGIKRNL  
FOKEGHPICSLRQLLEQFOKPEMNNVOKESPIVSVETNFDLSLGFPPKTHSVRSKSDTY  
VYNTKICLRTHSAHQPEFSLAKGFKNGFLITADVYRDEVDSSHVPIPHOMEG  
ALVWNRNDTAKMNDLKVSLSKSWDDTETALETNDQDQDYSLETKEFVSQHL  
KDTLTTVTHDLVSLAPSISGAEQVRYRTYDSFPFKPSQLEIDMKKWLILCGG  
VVQDRLLKGLANNYIGWAFGIGLERLAMLYGIPDIRLFWSLDERFSKQFIPNKIST  
FKFSPKACFKDIAFWINDFNPNDFEIIIRDVQDMVESVNLIDQYTKSGKTSLC  
YRVNYSRSLRNEIDKLOELRRNVANSRLVELR"  
complement(3184. .3402)  
/gene="SPCC736.03c"  
/note="Match to PF01409 tRNA-synt\_2d, tRNA synthetases  
class II (F) Score 80.50"  
misc\_feature  
complement(3208. .3237)  
/gene="SPCC736.03c"  
/note="PS00339 Aminoacyl-transfer RNA synthetases  
class-IIs signature 2"  
misc\_feature  
complement(3646. .3978)  
/gene="SPCC736.03c"  
/note="Match to PF01409 tRNA-synt\_2d, tRNA synthetases  
class II (F) Score 29.47"  
misc\_feature  
complement(3649. .3702)  
/gene="SPCC736.03c"  
/note="PS00179 Aminoacyl-transfer RNA synthetases  
class-IIs signature 1"  
gene  
complement(5197. .6324)  
/gene="gmal2"  
/note="synonym: SPCC736.04c"  
CDS  
complement(5197. .6324)  
/gene="gmal2"  
/EC\_number="2.4.1.-"  
/note="galactosyltransferase;  
alpha-1,2-galactosyltransferase (PMID 95003210);  
GNA12/MNN10 family; localization golgi (PMID 95003210);  
involved in galactosylation of cell surface glycoproteins  
O-linked oligosaccharides (PMID 9839953); predicted  
N-terminal signal sequence; similar to S. pombe gmh1 and  
gmh2 and gmh3 and SPBC1289.13C and SPAC5H10.13C and  
SPAC637.06 and SPBC82.17 (paralogues); similar to S.  
cerevisiae MNN10 and MNN11"  
/codon\_start=1  
/protein\_id="CAAL19268.1"  
/db\_xref="GI:3169074"  
/db\_xref="GOA:O09174"  
/db\_xref="InterPro:IPR008630"  
/translation="MRFAPYLISAVVITIIIGGAWTSMADTKLOTWKKEIIDQHTS  
TWPVWSSVTSTQDTLRVTISEVSVTATLTFTATFTVTSVVALATTDRHPDNS  
KIVILMGSNFQDANSPLHPPAQSIKNNREYAEHGYKFEFLDADAYASRVTHLMP

gene	WVKVPLQDTMKKYPDAEHWIWLHDLDALVMKDLNVVDHVLKHDRLNTILTRAEYKS GAGIPADGFFPKDQADKDVHFIISQDFNGINAGSLFIRNSEVGRWIVDLWFELIYLD HIQGYAQQAFSHVFPHPQYIRKHGVVPLKAINAYDFDDNIWGDGLCIHFAGCN YFKNCPKFLKYAQILSSQGSMDMSAQEKDHQNLKPPS" join(9089..9227,9436..9807,9859..9981,10026..10054) /gene="SPCC736.05" /note="synonym: wtf7" CDS join(9089..9227,9436..9807,9859..9981,10026..10054) /gene="SPCC736.05" /note="wtf element" /codon_start=1 /protein_id="CAAL19269.1" /db_xref="GI:4160568" /db_xref="GOA:Q72915" /db_xref="InterPro:IPR004982" /db_xref="UniProtKB/Swiss-Prot:Q72915" /translation="MSGSIAPIDESADELSVHSGNDNEIDLEKGLLPKCNKTNGGTTT CSEPPHYSDAVVGVENNIYMRTPFKLPFAKSGVTNSNVVFMKTYENNRHMT FRLRPPAKNGVTVGKLAQSLFLLPFNFIFPACLFPRKASFTDFSLGWLIFGICWL TCFLSFLIYAHESWTKPARSQPSLLIFGLLPFGIVTMVVFALYMRREVFLSL DK" misc_feature 9228..9233 /gene="SPCC736.05" /note="gtaagt, splice donor sequence" 9413..9435 misc_feature 9413..9435 /gene="SPCC736.05" /note="cctaattttttttttttttttcag, splice branch and acceptor" 9808..9813 misc_feature 9808..9813 /gene="SPCC736.05" /note="gtaagt, splice donor sequence" 9847..9858 misc_feature 9847..9858 /gene="SPCC736.05" /note="ttaaacttaag, splice branch and acceptor" 9982..9987	
Query Match	2.33; Score 74.8; DB 4; Length 44042; Best Local Similarity 48.8; Pred. No. 8e-10; Matches 202; Conservative 0; Mismatches 212; Indels 0; Gaps 0;	
QY	2675	TATGAAGAAGCAAGAAACTGGAATGTGAACCCATCATCCGGAACAACTGTGGATAA 2734
Db	25076	TAAGAACAAATCCGACGAGATAGAAATGGAATCCCTCCCTGGTACATTAATTGAAA 25135
QY	2735	ACTTATCGTTTCGAAATACAAATTCGATTTTCTTGGCATCTCATCATGTTGCTCTGG 2794
Db	25136	GCATGTTACTCACCCCTATCAGTACGATTTTACCTTATATCCACCCACTCTCTGCAAG 25195
QY	2795	TACATCTCGTCCAGGACATTCACATGTTATGTATGATGAGCAATGAAGCAAGATGA 2854
Db	25196	CGTTTCGGTCCCAAGTTCACTACACGGTTTTGCACGACGAAATTCAGATGCCCGGACCA 25255
QY	2855	AGTCTATAAATGACCTACGAGCTTGCTTTTCTCTGCTAGATGTGCAAAACCCATCTC 2914
Db	25256	ATTCGAACACTGTGTTTAAATCTTGTGATGTTTATGCAAGACCAAGTCTGTCTTC 25315
QY	2915	GTTGCGCTGTTCCGGTTCATTATGCTCATTTATCATGTGAAAAAGCAAGAGCTTTATCG 2974
Db	25316	CTTGGTTCACCTGTTTATTATGCGCATTGGTTAGTAATTCGCGGTATCAAGAGCT 25375
QY	2975	AACTTACAGGAAACATTCATCGGTGACTATGCAAGCCAGGACTGCACACGAATGGA 3034
Db	25376	GACAGCGATGATACATTTTGTGTGACATCAGAACATCAATGGATCAAGAAGTAAAGCC 25435
QY	3035	ACATTTTCTCCAACTAAACGTGAAGTACCTCGGATGTCTGTCGCAATACATTT 3088
Db	25436	GCCTTTTGTCTCTATCAAGTAAATTTGAAACAAAGATGTGGTATATGTTAATGTTT 25489
RESULT 6		
DQ178241		
LOCUS	4593 bp mRNA linear INV 19-SEP-2005	
DEFINITION	Caenorhabditis elegans argonaute-like (alg-1) mRNA, exons, and	

ACCESSION	DQ178241
VERSION	GI:75288353
KEYWORDS	Worm Transcriptome Project.
SOURCE	Caenorhabditis elegans
ORGANISM	Caenorhabditis elegans
REFERENCE	Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS	Grishok,A., Pasquinielli,A.E., Conte,D., Li,N., Parrish,S., Ha,I., Baillie,D.L., Fire,A., Ruvkun,G. and Mello,C.C.
TITLE	Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing
JOURNAL	Cell 106 (1), 23-34 (2001)
PUBMED	11461699
REFERENCE	2 (bases 1 to 4583)
AUTHORS	Morel,J.B., Godon,C., Mourrain,P., Beclin,C., Boutet,S., Feuerbach,P., Proux,F. and Vaucheret,H.
TITLE	Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance
JOURNAL	Plant Cell 14 (3), 629-639 (2002)
PUBMED	11910010
REFERENCE	3 (bases 1 to 4583)
AUTHORS	Kamath,R.S., Fraser,A.G., Dong,Y., Poulin,G., Durbin,R., Gotta,M., Kanapin,A., Le Bot,N., Moreno,S., Sohrmann,M., Welchman,D.P., Zipperlen,P. and Ahringer,J.
TITLE	Systematic functional analysis of the Caenorhabditis elegans genome using RNAi
JOURNAL	Nature 421 (6920), 231-237 (2003)
PUBMED	12529635
REFERENCE	4 (bases 1 to 4583)
AUTHORS	Simmer,F., Moorman,C., van der Linden,A.M., Kuijk,E., van den Bergher,P.V., Kamath,R.S., Fraser,A.G., Ahringer,J. and Plasterk,R.H.
TITLE	Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions
JOURNAL	PLoS Biol. 1 (1), E12 (2003)
PUBMED	14551910
REFERENCE	5 (bases 1 to 4583)
AUTHORS	Li,S., Armstrong,C.M., Bertin,N., Ge,H., Milestein,S., Boxem,M., Vidalain,P.O., Han,J.D., Chesneau,A., Hao,T., Goldberg,D.S., Li,N., Martiner,M., Rual,J.F., Lamesch,P., Xu,L., Tewari,M., Wong,S.L., Zhang,L.V., Berri,G.F., Jacotot,L., Vaglio,P., Reboul,J., Hirose,D., Kishikawa,T., Li,Q., Gabel,H.W., Glewa,A., Baumgartner,B., Rose,D.J., Yu,H., Bosak,S., Sequeria,R., Fraser,A., Baumgartner,B., Saxton,W.M., Strome,S., Van Den Heuvel,S., Piano,P., Vandenhaute,J., Sartet,C., Gerstein,M., Doucette-Stamm,L., Gunsalus,K.C., Harper,J.W., Cusick,M.E., Roth,F.P., Hill,D.E. and Vidal,M.
TITLE	A map of the interactome network of the metazoan C. elegans
JOURNAL	Science 303 (5657), 540-543 (2004)
PUBMED	14704431
REFERENCE	6 (bases 1 to 4583)
AUTHORS	Grishok,A., Sinskey,J.L. and Sharp,P.A.
TITLE	Transcriptional silencing of a transgene by RNAi in the soma of C. elegans
JOURNAL	Genes Dev. 19 (6), 683-696 (2005)
PUBMED	15741313
REFERENCE	7 (bases 1 to 4583)
AUTHORS	Ding,L., Spencer,A., Morita,K. and Han,M.
TITLE	The Developmental Timing Regulator AIM-1 Interacts with miRNAs and May Target the Argonaute Protein ALG-1 to Cytoplasmic P Bodies in C. elegans
JOURNAL	Mol. Cell 19 (4), 437-447 (2005)
PUBMED	16109369
REFERENCE	8 (bases 1 to 4583)
AUTHORS	Kohara,Y., Shin-i,T., Suzuki,Y., Sugano,S., Thierry-Mieg,D. and Thierry-Mieg,J.
TITLE	The Caenorhabditis elegans transcriptome project, a complementary view of the genome
JOURNAL	Unpublished
REFERENCE	9 (bases 1 to 4583)



AUTHORS	Kohara,Y., Shin-I,T., Potdevin,M., Thierry-Mieg,D. and Thierry-Mieg,J.
TITLE	Direct Submission
JOURNAL	Submitted (25-AUG-2005) Genome Biology Laboratory, Center For Genetic Resource Information, National Institute of Genetics, Mishima 411-8540, Japan
COMMENT	Summary: This essential gene alg-1 is also known in WormGenes/AceView by its positional name XO573, in Wormbase by its cosmid number name F48F7.1, in NextDB, the Nematode expression pattern database, as CEYK1605.  Expression level and number of variants According to the Worm Transcriptome Project, this gene is expressed at high level, 2.3 times the average gene in this release, at all stages of development [Kohara cDNAs], mainly in the germline of the developing and mature gonad, from L1 to adult, and in embryos [in situ hybridization to RNAs, see NextDB]. The protein [from a GFP translational fusion] is seen in the developing reproductive system (vulva, uterus), the pharynx, muscle, epidermis, head and tail neurons (expression in the germline cannot be observed in this experimental setting) [Johnsen R Chen L Mah A McKay S Tang E Tu D Zhao Z Baillie DL Newbury R Anastas D Viveiros R Moerman DG, as part of the BC C. elegans Gene Expression Consortium, 2005]. The sequence of this gene is defined by 26 cDNA clones and produces one transcript with one COOH-complete product. [Aug 2005, see wormgenes.org] The alg-1 gene is one of 27 genes (at least 23 known to be transcribed) with both a PAZ and a Piwi domain. The single variant from this gene has at least three distant polyadenylation sites, yielding 3' UTRs around 420 bp, 1210 bp and 1550 bp. These UTRs are remarkably long and provide ample space for regulative interactions.  mRNA Summary This mRNA is 4583 bp long, but it may be incomplete at the 5' end. The premessenger has 6 exons and covers 4.99 kb on the WS140 genome. The protein has 1010 aa (111.7 kDa, pI 9.2), contains one Argonaute and Dicer protein, PAZ domain and one Stem cell self-renewal protein Piwi domain. It also contains a 2nd peroxisomal domain [Psport2]. Taxblast results (threshold 10 <sup>-3</sup> ) track ancestors down to Eukaryota.  Product Molecular properties: [Bartel 2004] proposes localisation of ALG-1 and DCR-1 to the microRNA induced silencing complex (miRISC).  Phenotype Phenotypes and affected processes are: clear, translucent appearance, protruding vulva, uncoordinated locomotion, heterochronic, paralysed, small body size, slow growth, embryonic lethal.  [Grishok et al, 2001] alg-1 inactivation causes heterochronic phenotypes similar to lin-4 and let-7. RNA interference results [J. Ahringer 2003] Clear, uncoordinated, protruding vulva (by feeding genomic PCR product JA: F48F7.1). [Simmer et al, 2003] confirmed and extended Ahringer's RNAi phenotypic observations in the RNAi hypersensitive strain rrf-3: Gro, larval lethal, small adults, paralysed unc. [Rual et al, 2004] observed embryonic lethality in their RNAi experiment, using an interactome cDNA-like clone. Three knock out alleles, with identified non-overlapping deletions or rearrangements, have been isolated and are available from the CGC or from Shohel Mitani, Japan: gk214 (slightly small; www.celeganskoconsortium.omrf.org; strain VC446), tm492 (579 bp deletion, lethal or sterile, S. Mitani) and tm369 (a 774 bp deletion, lethal or sterile, S. Mitani). [Grishok et al, 2005] Transcriptional silencing (rather than the usual post-transcriptional silencing) of a transgene by RNAi occurs in the soma and is dependent on the PAZ-Piwi protein Alg-1 and on the Hpi homolog Hpi-2. This inhibition is accompanied by a decrease in the acetylation of histones associated with the transgene. This transcriptional silencing in the soma can be distinguished from transgene silencing in the germline by its inability to be
	transmitted across generations and its dependence on the rde-1 gene. [Ding et al, 2005] observed that an alg-1 deletion mutant displayed a retarded heterochronic phenotype in seam cell development with 94% defects, higher than in ain-1 mutants, at 40% defects. The double mutant has the same phenotype as alg-1 (93% defects). consistent with the two genes being in the same pathway and AIN-1 binds to ALG-1.
	Interactions This gene interacts with: gene lin-4: [Grishok et al, 2001] dcr-1, alg-1 and alg-2 are necessary for the maturation of lin-4. gene let-7: [Grishok et al, 2001] dcr-1, alg-1 and alg-2 are necessary for the maturation of let-7. gene hpi-2: [Grishok et al, 2005]. gene rde-1: [Grishok et al, 2005]. protein AIN-1: [Ding et al, 2005] showed that AIN-1 interacts with ALG-1 in vivo and in vitro, by coimmunoprecipitation using the GFP-tagged functional fusion protein in vivo, followed by mass spectrometry identification: AIN-1 is part of a protein complex containing both Dicer and Argonaute. Adding DNase I and RNase A to the worm lysate did not alter the binding pattern of ALG-1 and DCR-1 to AIN-1, excluding the possibility that nucleic acids mediate the binding among these proteins. This group shows that AIN-1 targets ALG-1 to cytoplasmic foci that are likely P bodies. Location/Qualifiers 1..4583 /organism="Caenorhabditis elegans" /mol_type="mRNA" /db_xref="taxon:6239" /chromosome="X" /map="X:14.41 cM (interpolated genetic position)" /clone="yk759f4, yk60e5, yk375c8, yk889c6, yk1013a7, yk1164h8" 1..4583 /gene="alg-1" /locus_tag="XO573" /note="Caenorhabditis elegans essential gene alg-1, encoding argonaute-like gene, required for maturation of microRNAs, and also RNAi-induced transcriptional gene silencing; regulates developmental timing; mutants are heterochronic; targeted to the miRISC complex by AIN-1" /map="X:14.41 (interpolated genetic position)" /db_xref="AceView/WormGenes:alg-1" /db_xref="NextDB:CELK01605" <1..3034 /gene="alg-1" /locus_tag="XO573" /note="required for maturation of microRNAs, and also RNAi-induced transcriptional gene silencing; regulates developmental timing; mutants are heterochronic; targeted
	Query Match 2.1%; Score 69; DB 13; Length 4583; Best Local Similarity 56.3%; Pred. No. 4.2e-08; Matches 129; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY	2716 CCGGAACAACCTGTGGATAAATTATTCGTTTCGAAATACAAATTCGATTTTCTTGGCAT 2775 
Db	2649 CAGGAACACTCTCGATGTAGGAATTACCCACCACCGAGTTTGACTTCTACTTGTGCT 2708 
QY	2776 CTCATCATGTCCTTGGTATCATCTCGTCCAGGACATTACACTGTATGTATGACGATA 2835 
Db	2709 CTCATGCTGTATTTCAGGAACATCTCGTCCATCCATTACCATGTTCTTTGGGATGACA 2768 
QY	2836 AAGGAATGAGCCCAAGATGAAGTCTATATAAAATGACCTACCGACTGCTTTTCTCTGCTGA 2895 
Db	2769 ACAATCTGACAGCCGACGAACCTCAACAGCTCACATATCAAAATGTCCATACCTACGTCA 2828 
QY	2896 GATGCGAAACCCATCTCGTTCCTGTTCCGGTTTCATTATGCTCATTT 2944 
Db	2829 GATGCACACGATCCGTTTCAATTCCAGCCGCGCATATTATGCTCATTT 2877 
FEATURES	source
	gene
	CDS



RESULT 7  
CEP48F7

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## CONSTRM

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CEP48F7 29820 bp DNA linear INV 09-AUG-2005  
Caenorhabditis elegans Cosmid F48F7, complete sequence.

Z69661

Z69661.1 GI:1200275

HTG; Syntaxin like.

Caenorhabditis elegans

Caenorhabditis elegans

Bukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 29820)

C. elegans Sequencing Consortium

Genome sequence of the nematode C. elegans: a platform for

investigating biology

Science 282 (5396), 2012-2018 (1998)

9851916

2 (bases 1 to 29820)

Coles, L.

Direct Submission

Submitted (24-FEB-1996) Nematode Sequencing Project, Sanger

Institute, Hinxton, Cambridge CB10 1SA, England and Department of

Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:

wormsanger.ac.uk

Coding sequences below are predicted from computer analysis, using  
predictions from GeneFinder (P. Green, U. Washington), and other  
available information.

Current sequence finishing criteria for the C. elegans genome

sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a

dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of

the specified clone. It may be shorter because we only sequence

overlapping sections once, or longer because we arrange for a small

overlap between neighbouring submissions.

For a graphical representation of this sequence and its analysis

see: - [http://www.wormbase.org/perl/ace/elegans/seq/sequence?](http://www.wormbase.org/perl/ace/elegans/seq/sequence?name=F48F7;class=Sequence)

names=F48F7;class=Sequence

IMPORTANT: This sequence is not the entire insert of clone F48F7.

It may be shorter because we only sequence overlapping sections

once, or longer because we arrange for a small overlap between

neighbouring submissions.

The true left end of clone F48F7 is at 1 in this sequence. The true

right end of clone F48F7 is at 12044 in

sequence Z78539.

The true left end of clone C31E10 is at 29717 in this sequence. The

true right end of clone K02B9 is at 5215 in this sequence. The

start of this sequence (1..104) overlaps with the end of sequence

Z69663.

The end of this sequence (29717..29820) overlaps with the start of

sequence Z78539.

Location/Qualifiers

1..29820

/organism="Caenorhabditis elegans"

/mol\_type="genomic DNA"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="X"

/clone="F48F7"

join(Z69663.1:30546..30755,112..284,431..601,655..2117,

2173..3125,3178..3216)

/genes="alg-1"

/locus\_tag="F48F7.1"

join(Z69663.1:30546..30755,112..284,431..601,655..2117,

2173..3125,3178..3216)

/genes="alg-1"

/locus\_tag="F48F7.1"

/standard\_name="F48F7.1"

/notes="C. elegans ALG-1 protein; contains similarity to

Pfam domains PF02170 (PAZ domain), PF02171 (Piwi domain)"

```

/codon_start=1
/product="Hypothetical protein F48F7.1"
/protein_id="CAA93496.2"
/db_xref="GI:22265872"
/db_xref="InterPro:IPR003100"
/db_xref="InterPro:IPR003165"
/db_xref="UniProtKB/TrEMBL:Q20578"
/translation="MSGGQYLFGVNMNSTIQQQPSATSSFLSPGSPISSTSTSQVVP
TSGATQPPFSAQAASATALNDLEIFNSPPTQPTQFSDVPORQAGSLAPGVPIGN
TSVSGIPANTLGGGLPGGAPGOLPGGNQSGIOFQCRPRRNVHVEGRSILLRANHFV
RIPGRTIOHYQDVTPDKCPRVNRREIISCLISAPSKYFTNIRPVYQKRNMYTREPL
PIGRMDPVTLPDGSNAVERQPSVSLKWQVSLSDLEDAWEGRVRQVPEAQQAMD
VIIRHLPSLKTPTVGRSFFSPFVPMNAGSCPCPQAGAVAGSAGSAGSAGSAGSAGS
GGREVFHGFQHSVRSPQKMLNIDVSATAFYRSMVEVIEFIAEVLPLVQALAEERA
LSDAQRFKTEIRGLKIEITHCGQMRKRVNNTVRPAQTQTFPQLEGTQCTIECT
VAKYFYDKYIQLKPHLPCLOVGOEKHTYLPPEVCNIPVGCRCIKKLTDVQSTMI
KATARSAPERREISNLVRKAEPFADFAHEFGITINPAMTEVKGVLSPKLYGGR
TRATLNPQGWDMRGKQFHTGIDVRWAJACPAQQQHVXENDLRMTNQLRISNDRA
GMPIVGNPCFKYAVGEVPEPMFKYKLYNQYSGIQLVVVLPKPTPYAEVVRVGDTV
LGATQCVAKNAIRTPPTLSNLKLNKQYKLVGVSILLPNVRPFIENEPVIFGCD
ITHPPAGDSRKPISAAVVGSMDAHPSRYAATVRQHRQBIISLTYMRELLVQFVR
NTRPKPARIVVYRDVSGEQFENVLOYELRAIIRACMMLERGVOPGTTTAVQKRRHT
RLFAVDKQVQKAYNIPGTTVDVGTHTPEPDLCLSHAGIGCTSPRSHYHVMDD
NNTLDELQLTQYMCHTYVRCRSVISPAPATYAHLVAFRARTHLVDREHDSGEGSQ
PGSTSDTLTUSNWARVQVHPDANNVYFA"
join(8676..8702,9134..9220,9295..9396,9446..9545,
10008..10219,10265..10417,10465..10581,10842..10943)
/genes="syn-2"
/locus_tag="F48F7.2"
join(8676..8702,9134..9220,9295..9396,9446..9545,
10008..10219,10265..10417,10465..10581,10842..10943)
/genes="syn-2"
/locus_tag="F48F7.2"
/standard_name="F48F7.2"
/notes="C. elegans SYN-2 protein; contains similarity to
Pfam domains PF05739 (), PF00804 (Syntaxin)"
/codon_start=1
/product="Hypothetical protein F48F7.2"
/protein_id="CAA93492.4"
/db_xref="GI:54110860"
/db_xref="GOA:Q20574"
/db_xref="HSSP:JUTH"
/db_xref="InterPro:IPR000727"
/db_xref="InterPro:IPR006011"
/db_xref="InterPro:IPR006012"
/db_xref="UniProtKB/Swiss-Prot:Q20574"
/translation="MEDRLNEFSQSVTDRFDEVELSPARPSPSAEYVDRFEFERNAL
ASVRGIEKLRRDQHQVLTATADPRDKNILENQITIRRTGDLRKLVRQADDFFLE
FTQGVSVTEKMRQQLKLLKLNKLNLFNETHODYKRSVSRVRRQLQVGRQDL
TDEDINRMENSGSEQLPFRVNPVLSGQAAYEDVKRGEIKDLNNIAMLKEEFL
DLQHLTEAQDEMVTNIDNNVENGLEQVKGQSANVKTAVEYKKSAMRKKICVAAILIT
LILITLITAILAVLSRGNNKK"
join(12943..12981,13029..13154,13201..13313,13361..13565,
13887..14059,14104..14209)
/locus_tag="F48F7.3"
join(12943..12981,13029..13154,13201..13313,13361..13565,
13887..14059,14104..14209)
/locus_tag="F48F7.3"
/standard_name="F48F7.3"
/notes="contains similarity to Pfam domain PF01762
(galactosyltransferase)"
/codon_start=1
/product="Hypothetical protein F48F7.3"
/protein_id="CAA93493.1"
/db_xref="GI:3877296"
/db_xref="GOA:Q20575"
/db_xref="InterPro:IPR002659"
/db_xref="UniProtKB/TrEMBL:Q20575"
/translation="MQRENFPGAWLNENNAVMYFIVGTIDVIDIDDEIKTYEDIIQVD
INENYHNTYKAIFWINEFKRGHPKLLKLLKDDVDHIDMIGLQFLIKRYSVMDDFMA
CVISNGVQVRNNTSKWLSKKDYKPNLGTVCQGMVYFISGNLLPLHGNIEKQFL
WNDDVTRSLVGDYKISYISLEQHSISIPNTVREYVSYSLMAIKTRKWTTFIAHFRFPQ
KPFMRRRRKIWANIITEINNCSNILRKARINFVSAY"

```

gene	complement(join(14326..14392,14451..15271,15331..15760,15814..16025,16084..16308,16370..16450,16726..16893,16942..17165,17346..17558,17741..18077)) /gene="pqn-39" /locus_tag="F48F7.4" complement(join(14326..14392,14451..15271,15331..15760,15814..16025,16084..16308,16370..16450,16726..16893,16942..17165,17346..17558,17741..18077)) /gene="pqn-39" /locus_tag="F48F7.4" /standard_names="F48F7.4" /note="C. elegans PQN-39 protein; contains similarity to Saccharomyces cerevisiae Substrate of the HupB ubiquitin-like protein that localizes to the shmoo tip (mating projection); mutants are defective for mating projection formation, thereby implicating HupB in polarized cell morphogenesis; SGD:YDL223C"
CDS	/codon_start=1 /product="Hypothetical protein F48F7.4" /protein_id="CAA93494.1" /db_xref="GI:3877297" /db_xref="UniProtKB/TrEMBL:Q20576" /translation="MLCFSCRSLFQFNFLLSILFLFTNSHLLTISLSSSST PWCFLQRRRRASVERNVCLSWACGRORNMVKGPTTALLIVVISDGRGVDDGN EGSRGIAFSRFLVFLCLSLSQPFILQSOYISVFCVQFQFANQILQLSLGLKH FKLYSTFLFRCSKLYQYPACNSSFLVMSAKODLAPAPVSHADTGLQLPP KQVHPNTHGPIIYQSYAPVAHYQGPQVWNPQSTVHVYQAPDASSQEN KQVSNHYGTHYQAQPMNDMLRPQQATYVQYVTPPQYPAQQTQTFISYA TTFVANDYTHPNHMRGAGRDYHVAQSSVGMVFVFPVATGVNMQAHA SASYATNVSSTSIPEMKDPESDNIQYDQGRPTSLNGENWTRDAEDHLSPTTV NHPNVNRPQNFVEKESQSTTRMSKGRNKGKRLTDFDLHAKWQKS RVNENVGATSPATTVAAPSNVANSSTALTPRNSQCKETSRYRNEGSNNQ NRNTKRVGANGHMYNNEGHSAGHEQFQQQQQYVYNGATGERQRPRNDY SKNRVYNNHNSNNRSPYQAQNNQPTQONHPREGAPENYQOSYRGNRRG GAYRGYHQHQNQLNPQPLNQYPPAQIMGPQWATMNAFQVANTGTLPPG VORIPLPGFQWPRVHODIKFQNTWLDATSCPLPAAVMGIDTRFGRGMRGRG VNVYVRGVRGTVGVAAPRYIGRPVETPATPKSEOSSSTEVDMAKDDERSVQ DRASATESVETPADQGTQEQNKQSDSQNGQEPVAKANBEIILLPANTVQSPD AAAEKKEVEST" join(24169..24244,24310..24416,24468..24656,24863..25019,25077..25429) /locus_tag="F48F7.5" join(24169..24244,24310..24416,24468..24656,24863..25019,25077..25429) /locus_tag="F48F7.5" /standard_names="F48F7.5" /notes="contains similarity to Pfam domain PF00096 (Zinc finger, C2H2 type)" /codon_start=1 /product="Hypothetical protein F48F7.5" /protein_id="CAA93490.1" /db_xref="GI:3877293" /db_xref="GOA:Q20572" /db_xref="InterPro:IPR007087" /db_xref="UniProtKB/TrEMBL:Q20572" /translations="MADIPKRIFDINQQLKIETCGTVMDMYKLCQCKDEKENTENS FAIHDIUKNRERDHYVQVQHARLLMSQNLAMFCGASNQEDLKCGLCKEQPEN
Query Match	2.1%; Score 69; DB 13; Length 29820;
Best Local Similarity	56.3%; Pred. No. 3.9e-08;
Matches 129; Conservative	0; Mismatches 100; Indels 0; Gaps 0;
QY	2716 CCGGAACAACCTGGGATAAATCTATCGTTTCGGAATAACAATTCGATTTTCTTGGCAT 2775
Db	 2779 CAGGAACACTGTCGATGTAGGAATTACCCACCAACCGAGTTTGACTTCTACTTGTGCT 2838
QY	2776 CTCATCATGTGTCCTTGGTACATCTCGTCCAGACATTACACTGTTATGTATGACGATA 2835
Db	 2839 CTCATGCTGTTATCAAGGAACATCTCGTCCATCCCATCAACATGTTCTTTGGGATGACA 2898
QY	2836 AAGGAATGACCGCAAGATGAAGTCTATAAATGACCTACCGACTTGCTTTTCTCTCTGCTA 2895
Db	 2899 ACAATCTGACAGCGGACGAATCTCAACAGCTCACATATCAAAATGTGCCATACCTACGTCA 2958
QY	2896 GATGTGCAAAACCCATCTCGTTGCGTGTTCGGTTTCGGTTTCATTATCTCATTTT 2944
Db	 2959 GATGCACACGATCCGTTTCAATTCACGCGCCAGCATATTATGCTCATTT 3007
RESULT 8	
CEZK757/c	CEZK757 25850 bp DNA linear INV 09-AUG-2005
LOCUS	Caenorhabditis elegans Cosmid ZK757, complete sequence.
DEFINITION	CEZK757
ACCESSION	Z29121
VERSION	Z29121.1 GI:438366
KEYWORDS	HTG.
SOURCE	Caenorhabditis elegans
ORGANISM	Caenorhabditis elegans
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE	1 (bases 1 to 25850)
AUTHORS	C. elegans Sequencing Consortium
CONSTRM	Genome sequence of the nematode C. elegans: a platform for
TITLE	Investigating Biology
JOURNAL	Science 282 (5396), 2012-2018 (1998)
PUBMED	9851916
REFERENCE	2 (bases 1 to 25850)
AUTHORS	Thomas, K.
TITLE	Direct Submission
JOURNAL	Submitted (02-JAN-1994) Nematode Sequencing Project, Sanger
	Institute, Hinxton, Cambridge CB10 1SA, England and Department of
	Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
	worm@sanger.ac.uk
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. For a graphical representation of this sequence and its analysis see: <a href="http://www.wormbase.org/perl/ace/elegans/seq/sequence?name=ZK757;class=Sequence">http://www.wormbase.org/perl/ace/elegans/seq/sequence?</a> name=ZK757;class=Sequence IMPORTANT: This sequence is not the entire insert of clone ZK757. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The start of this sequence (1..414) overlaps with the end of sequence Z22180. The end of this sequence (25695..25850) overlaps with the start of sequence Z30215.
FEATURES	Location/Qualifiers
Source	1..25850 /organism="Caenorhabditis elegans" /mol_type="genomic DNA" /strain="Bristol N2" /db_xref="taxon:6239" /chromosome="III" /clone="ZK757" join(complement(3000..3053), complement(1928..2009), complement(1666..1787), complement(1415..1623), complement(1285..1362), complement(1038..1128), complement(888..950), complement(684..798), complement(113..239), complement(222180..1:32171..32424), complement(222180..1:31663..31679)) /locus_tag="ZK757.4a" join(complement(3000..3053), complement(1928..2009), complement(1666..1787), complement(1415..1623), complement(1285..1362), complement(1038..1128), complement(888..950), complement(684..798),
gene	
CDS	

```

complement (113..239), complement (222180.1:32171..32424),
complement (222180.1:31663..31679))
/locus_tag="ZK757.4a"
/standard_name="ZK757.4a"
/notes="contains similarity to Pfam domain PF01529 (DHHC
zinc finger domain)"
/codon_start=1
/product="Hypothetical protein ZK757.4a"
/protein_id="CAD54174.2"
/db_xref="GI:24817642"
/db_xref="GOA:Q810G4"
/db_xref="InterPro:IPR001594"
/db_xref="UniProtKB/Swiss-Prot:Q810G4"
/translations="MSCRDCTKGGCVYTFWLVFLPVVLVSLATGWIYATYELC
ILSDNPQRILYIFFYALLILFYTLRTVYTKWPKQKCYEGASATYSEVK
DEQLQFLSDIARERDITLLVRGDFHGRFCDCCKCIPDRSHSCMCKQKFDH
HCPVNNVNGFYKYIFLFLAYGFIICWIAATLPSFIDFWRHEYDMCKQVDSID
SVIQRLKHLTVLSNGRFLVLLFLSCMFSLSLFLFVHLVLTAKNRTTVESPA
PMIDGKADAFNHGIRANYREIFGSHPLVYFELPVPSISIGDCGKFPVNDTMSAAG
NQPVENGVPNGQSHAPPIAQHHINLYSEQSSASEKEIKSVDDBITERSQVST
TAPSHDSVTA"
join (complement (2427..2486), complement (1928..2009),
complement (1666..1787), complement (1415..1623),
complement (1285..1362), complement (1038..1128),
complement (888..950), complement (684..798),
complement (113..239), complement (222180.1:32171..32424),
complement (222180.1:31663..31679))
/locus_tag="ZK757.4b"
/locus (complement (2427..2486), complement (1928..2009),
complement (1666..1787), complement (1415..1623),
complement (1285..1362), complement (1038..1128),
complement (888..950), complement (684..798),
complement (113..239), complement (222180.1:32171..32424),
complement (222180.1:31663..31679))
/standard_name="ZK757.4b"
/codon_start=1
/product="Hypothetical protein ZK757.4b"
/protein_id="CAH19096.1"
/db_xref="GI:51587415"
/db_xref="GOA:Q65ZC4"
/db_xref="InterPro:IPR001594"
/db_xref="UniProtKB/TrEMBL:Q65ZC4"
/translations="MCDVFNKIKLDDILKLVFLPVVLVSLATGWIYATYVE
LCILSDNPQRILYIFFYALLILFYTLRTVYTKWPKQKCYEGASATYSEV
KDERQLQFLSDIARERDITLLVRGDFHGRFCDCCKCIPDRSHSCMCKQKFLF
DHCPVNNVNGFYKYIFLFLAYGFIICWIAATLPSFIDFWRHEYDMCKQVDS
IDSVIQRLKHLTVLSNGRFLVLLFLSCMFSLSLFLFVHLVLTAKNRTTVESF
RAPMIDGKADAFNHGIRANYREIFGSHPLVYFELPVPSISIGDCGKFPVNDTMSA
AGNQPVENGVPNGQSHAPPIAQHHINLYSEQSSASEKEIKSVDDBITERSQVST
ATTAPSHDSVTA"
complement (join (3966..4183, 4237..4368, 4674..4764,
4849..4944, 5337..5432, 5479..5658, 5713..5818, 6332..6387))
/locus_tag="ZK757.1"
complement (join (3966..4183, 4237..4368, 4674..4764,
4849..4944, 5337..5432, 5479..5658, 5713..5818, 6332..6387))
/locus_tag="ZK757.1"
/standard_name="ZK757.1"
/notes="contains similarity to Pfam domain PF06681 ()"
/codon_start=1
/product="Hypothetical protein ZK757.1"
/protein_id="CAA82387.3"
/db_xref="GI:24817641"
/db_xref="GOA:P34679"
/db_xref="InterPro:IPR010601"
/db_xref="UniProtKB/Swiss-Prot:P34679"
/translations="MIYNNISLNTITTTTLTYRDLTFPSVHGTCVFNFLFCIPFI
CRPLRLTKFTPTFFVLGTFVLSLFLFQVYLVVFLWSLPSVPRITIAVCLTKCIT
SSTTSQAQLPLAVAIYRPIVVRNKKMPSFWVVHVSISFIFFVIAILNPLGSGE
TNDQCAVLRFSKAMEARISLTGLNLFVAVINVAITYFKYDKRNDVHRRRVOLT
YMLQSMIPILVSIPLLVGSPDFYFGSLFELLWIKCKKISCFIHFRIYPSWLHVS
LVNNTISPSLADTNLNAFTSNNSPRVTLHLYFELFHRSNENIKFSY"
complement (join (9579..9810, 10116..10180, 10227..10370),

```

```

10695..10794, 11316..11422, 11470..11603, 12543..12645))
/locus_tag="ZK757.2"
complement (join (9579..9810, 10116..10180, 10227..10370,
10695..10794, 11316..11422, 11470..11603, 12543..12645))
/locus_tag="ZK757.2"
/standard_name="ZK757.2"
/notes="contains similarity to Pfam domain PF00782 (Dual
specificity phosphatase, catalytic domain)"
/codon_start=1
/product="Hypothetical protein ZK757.2"
/protein_id="CAA82388.2"
/db_xref="GI:30424335"
/db_xref="GOA:P34680"
/db_xref="InterPro:IPR000340"
/db_xref="InterPro:IPR000387"
/db_xref="UniProtKB/Swiss-Prot:P34680"
/translations="MPSRRILNREORQCSQIRPVLYVYSLAALSPRVLSRFCVCINLI
PGRSLAPPMKVVHILPLQNETDITLSPHWANYKEIEBARKGGRALLCMMGISRS
ATFGIATVMOYERKTLHDSYKAVOLARNIICPNVFPFQQLIDLEQLRGKVSCKIIEP
LPCKQPDVITWQELYDEMIMSQDHRSLASCNLSARSTNDTMSLRSLMNVNDTSR
KSLAPFLTHRPIGASPTLLVPSSSSSSVRGPIPLQRAHTEPPKEASILPKSALRDKS
KSGEKKKWRLSFHKQVV"
join (complement (230215.2:861..915),
complement (230215.2:201..373), complement (25676..25848),
complement (24734..25627), complement (24473..24689),
complement (23916..24423), complement (23604..23866),
complement (23434..23558), complement (22889..23386),
complement (22717..22841), complement (22579..22670))
/genes="tag-76"
/locus_tag="ZK757.3a"
join (complement (230215.2:861..915),
complement (230215.2:201..364), complement (25676..25848),
complement (24734..25627), complement (24473..24689),
complement (23916..24423), complement (23604..23866),
complement (23434..23558), complement (22889..23386),
complement (22717..22841), complement (22579..22670))
/genes="tag-76"
/locus_tag="ZK757.3a"
/standard_name="ZK757.3a"
/notes="contains similarity to Pfam domains PF02170 (PAZ
domain), PF02171 (PwI domain)"
/codon_start=1
/product="Hypothetical protein ZK757.3a"
/protein_id="CAA82389.1"
/db_xref="GI:3881773"
/db_xref="InterPro:IPR003100"
/db_xref="InterPro:IPR003165"
/translations="MSRRNATSFVNDNLTSSIGSGSMSPPISTRPASQASPLTS
NGSLSPVADQGSVSYNLDSPRDLSPILLSLACLNREVVARPCLGTIGRIQIPVK
SNFFPAMDLANPKMVTVOHVEIHHPCGCKLDKDEMIIFWKAVSDHNPINPKFALAY
DGAHQLYTVARLEFPDDQGSVLDCEASLPKONDRTRCAISIQNVGVLLENQRTRT
NLNLERVLTPQILDIDICRQSLCTPLLNKSNANPYTWKSCYRIPTAAGQALDLEGKRE
MWTGPFSSAHIASNRYRPLNIDVAHTAFYKTRITRVLFQFMDVLNERTSPNRNRPGR

```

```

Query Match      2.1%; Score 68.8; DB 13; Length 25850;
Best Local Similarity 55.4%; Pred. No. 4.4e-08;
Matches 133; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 2705 GAACCCATCATCGGAACAACCTGTGGATAACATTATCGTTTCGAAATACAAATTCGATTT 2764
      |||||
Db 23167 GAAGGTTCCACCTGGTACACAGTTCACACTGGAATCGTCTCTCCCGAAGGATTGATTT 23108
      |||||

QY 2765 TTTCCTGGCATCTCATCATCGGTGTCCTTGGTATACATCTCGTCCAGGACATTACACTGTAT 2824
      |||||

```

Db	23107	CTATCTGTGCTCTCACTACGAGTACAGGAACTTCTCGTCTCGCAATATCATGTTCT	23048
Qy	2825	GTATGACGATAAAGGAATGAGCAAGATGAAGTCTATAAAATGACCTACGGACTTGCTTT	2884
Db	23047	TCTGGATGAATCAAGTCTCACTGCTGATGAATCCAAAGCATCACCTACGGAATGTGCA	22988
Qy	2885	TCTCTCTGCTAGATGCGAAACCCATCTCGTTCCTGCTTCGGTTTCATTATGCTCATTT	2944
Db	22987	TACATATGGTGGTGTACTCGTTCGCTCCATTCCAACTCCAGTTTATTATGCTGATTT	22928
RESULT 9			
AF016682			
LOCUS	AF016682	33706 bp DNA linear	INV 22-SEP-2004
DEFINITION	Caenorhabditis elegans cosmid T07D3, complete sequence.		
ACCESSION	AF016682		
VERSION	AF016682.2	GI:49035142	
KEYWORDS	HTG.		
SOURCE	Caenorhabditis elegans		
ORGANISM	Caenorhabditis elegans		
REFERENCE	1 WormBase Consortium		
AUTHORS	Genome sequence of the nematode C. elegans: a platform for		
CONSTRM	investigating biology. The C. elegans Sequencing Consortium		
TITLE	Science 282 (5396), 2012-2018 (1998)		
JOURNAL	9851916		
PUBMED	2 (bases 1 to 33706)		
REFERENCE	Lamar, B., Wameley, P. and Twyman, B.		
AUTHORS	The sequence of C. elegans cosmid T07D3		
TITLE	Unpublished (2001)		
JOURNAL	3 (bases 1 to 33706)		
REFERENCE	Waterston, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (31-JUL-1997) Department of Genetics, Washington		
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
REFERENCE	Louis, MO 63110, USA		
AUTHORS	4 (bases 1 to 33706)		
TITLE	Waterston, R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (25-JUL-2001) Department of Genetics, Washington		
AUTHORS	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
TITLE	Louis, MO 63110, USA		
JOURNAL	5 (bases 1 to 33706)		
REFERENCE	Waterston, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (21-SEP-2001) Department of Genetics, Washington		
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
REFERENCE	Louis, MO 63110, USA		
AUTHORS	6 (bases 1 to 33706)		
TITLE	Waterston, R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (05-OCT-2001) Department of Genetics, Washington		
AUTHORS	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
TITLE	Louis, MO 63110, USA		
JOURNAL	7 (bases 1 to 33706)		
REFERENCE	Waterston, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (23-MAY-2002) Department of Genetics, Washington		
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
REFERENCE	Louis, MO 63110, USA		
AUTHORS	8 (bases 1 to 33706)		
TITLE	Waterston, R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (29-OCT-2002) Department of Genetics, Washington		
AUTHORS	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
TITLE	Louis, MO 63110, USA		
JOURNAL	9 (bases 1 to 33706)		
REFERENCE	Waterston, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (21-NOV-2002) Department of Genetics, Washington		
JOURNAL			

University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
 10 (bases 1 to 33706)  
 Waterston, R.

Direct Submission  
 Submitted (10-FEB-2003) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
 11 (bases 1 to 33706)  
 Wilson, R.

Direct Submission  
 Submitted (22-JUN-2004) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
 12 (bases 1 to 33706)

WormBase Consortium  
 Direct Submission  
 Submitted (22-SEP-2004) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
 On Jun 22, 2004 this sequence version replaced gi:2315759.  
 Submitted by:

Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1HQ, England  
 email: submissions@watson.wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this clone sequence and its analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=T07D3;class=Sequence>

#### NEIGHBORING CLONE INFORMATION

The 5' clone is Y46B2A, 200 bp overlap; the 3' clone is R07C3, 200 bp overlap. Actual start of this clone is at base position 1 of T07D3; actual end is at 2312 of R07C3.

#### NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara ([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C. elegans ORFome cloning project (<http://worldb.dfci.harvard.edu/>), similarity to other proteins from BlastX analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

#### FEATURES

Location/Qualifiers

1..33706  
 /organism="Caenorhabditis elegans"  
 /mol\_type="genomic DNA"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"

source

```

/chromosome="II"
/clon="T07D3"
750..6014
/gene="alg-2"
/locus_tag="T07D3.7"
join(750..804,1698..1800,2398..2571,2641..2981,3273..4334,
4566..5524,5976..6014)
/gene="alg-2"
/locus_tag="T07D3.7"
/standard_name="T07D3.7a"
/notes="contains similarity to Pfam domains PF02170 (PAZ
domain), PF02171 (Piwi domain); coded for by the following
C. elegans cDNAs: OSTF122A8.1, yk81c3.5,
yk109d4.5, yk160h5.5, yk175g11.5, yk185g2.5, yk199g3.5,
yk334f12.5, yk433a5.5, yk464f7.5, yk471e4.5, yk504h5.5,
yk515b9.5, yk519f3.5, yk564d8.5, yk592d7.5, yk609g12.5,
yk635g10.5, yk651b2.5, yk1044a12.3, yk1044a12.5,
yk1063d06.3, yk1063d06.5, yk1290a10.3, yk1290a10.5,
CEESW71R, yk1689c07.3, yk67f3.3, yk81a6.3, yk81c3.3,
yk109d4.3, yk160h5.3, yk334f12.3, yk433a5.3, yk515b9.3,
yk564d8.3, yk609g12.3, yk635g10.3, yk651b2.3, yk679a4.3"
/codon_start=1
/product="Argonaute (plant)-like gene protein 2, isoform
a"
/protein_id="AA866197.2"
/db_xref="GI:24418217"
/db_xref="WormBase:T07D3.7a"
/translations="MFPLPVHNGPRGLKLSIFMPGDSLTSSFPDGGAEFTSSSSQL
GSAHGAIGTKPAGQVQCPVRPNHGVGRSILLRANHPVRIPGGSVOHYOIVDPP
CKPRVNRVIVICLLSSRSKYFTNIRPVYDGRNMVTRPLPGTPEPNFVETLPGD
SAVERFSVTMWIGVQVCLSDADMEGRVQVPEAVQSIDVILRHLPSLTKTPVGR
SFTTPGVKPMQMGQVSKLGGREVFQGHOSVPSQWMMNLINDSATAFYRAMP
VIFVAEVLPEVQALAEARLSDAQRVFTGIRGLKIEITHCGAVRRKRYCVNTR
PQAOTFTPLQLEGTQIECTVAKYFDFKRYIQLPKPLQVQGEQKHLYLPPVVC
DIVPQGRCLKLTDVOTSMIKATARSAPEREREICKLVSKAELSDPAHAGITIN
PAMTEVGRVLSAPKLYGRRHATTALPNQGVDMRGQFHTGMEVVRWJAJACFAQ
SHVENDLRMFTQLRISTDAGNPILGTFMFCYASGVQVEPMPKYLKQTSALQL
IVVLPGKPIYAEVRVGVGTATQCVQAKNAIRTPQLTSLNLSLKNVNLVGVQV
SILLNPRRIEFYEFILGCDITHPAAGTRKPSIAAVGSDMAHPSRYAAVTRVQQ
HROEITITDLYVRELLQVYRNTFRKPARIVYRDGVSEGQLFNVLQVELRAIREAC
VLMESGQGITIAVQKQHRHLRAADQVQKAFNPFGTVDVGTHTPEPFF
LCSHAGIQTSRPSHVHLWDDNDLTADELQQLTYOMCHTVRCTSVSPAPAYAH
LVAFARYHLVDRDCHSGSEGSQPSGTSSEDTLLSNKAVQVHPSPSNWYFA"
join(1700..1800,2398..2571,2641..2981,3273..4334,
4566..5524,5976..6014)
/gene="alg-2"
/locus_tag="T07D3.7"
/standard_name="T07D3.7b"
/notes="contains similarity to Pfam domains PF02170 (PAZ
domain), PF02171 (Piwi domain); coded for by the following
C. elegans cDNAs: OSTF122A8.1, OSTR122A8.1, yk67f3.5,
yk81c3.5, yk109d4.5, yk160h5.5, yk175g11.5, yk185g2.5,
yk199g3.5, yk334f12.5, yk433a5.5, yk464f7.5, yk471e4.5,
yk504h5.5, yk515b9.5, yk519f3.5, yk532d3.5, yk564d8.5,
yk592d7.5, yk609g12.5, yk635g10.5, yk651b2.5, yk1044a12.3,
yk1044a12.5, yk1063d06.3, yk1063d06.5, yk1290a10.3,
yk1290a10.5, CEESW71R, yk1689c07.3"
/codon_start=1
/product="Argonaute (plant)-like gene protein 2, isoform
b"
/protein_id="AA038604.1"
/db_xref="GI:28275097"
/db_xref="WormBase:T07D3.7b"
/translations="MPGDSLTSSFPDGGAEFTSSSSQLGSAHGAIGTKPAGVQFQ
CPVRPNHGVGRSILLRANHPVRIPGGSVOHYOIVDPPDKPRVNRVIVICLLISF
SKYFTNIRVYDGRNMVTRPLPGTPEPNFVETLPGDSAVERFSVTMWIGVQVCL
SALDDAMEGRVQVPEAVQSIDVILRHLPSLKYTPVGRSFTTPGVKPMQMGHQS
2.1%; Score 68.2; DB 13; Length 33706;
Query Match
Best Local Similarity 55.2%; Pred. No. 6.5e-08;
Matches 133; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
2704 TGAACCATCATCCGGAACACTGTGTGATAAATTATCGTTTCGAATAACAATTCGATT 2763

```

```

Db 5160 TCAACATCCACAGGAACAACCTGTAGATGTGGAATCACGCATCCAAOGGAATTCGACT 5219
Qy 2764 TTTTCTTGGCATCTTCATCATGTGTCTTGTGTACATCTCGTCCAGGACATTACACTGTGTA 2823
Db 5220 TCTTCTCTCTCTCCACGCTGTATCCAGGAACATCCGACCATCTCCTACCTACCATGTGC 5279
Qy 2824 TGTATGACGATAAGGAATGAGCAAGATGAAGTCTATATAAATGACCTACGAGCTTGCTT 2883
Db 5280 TCTGGGATGACAATGATCTGACGCCGACGAGCTTCAACAGCTTGACCTTACCAAAATGTGTC 5339
Qy 2884 TTCTCTCTGTAGATGTCGAAACCCATCTCGTTGCTCTGTTCCGGTTCATTATCTGCTATT 2943
Db 5340 ATACCTATGTAAGATGACGAGGTCTGTGTGATCCCGCGCGCGGTACTATGCATC 5399
T 2944
5400 T 5400

RESULT 10
AJ719770
LOCUS AJ719770 4339 bp mRNA linear VRT 12-JAN-2005
DEFINITION Gallus gallus mRNA for hypothetical protein, clone 6f12.
ACCESSION AJ719770
VERSION AJ719770.1 GI:53129948
KEYWORDS ORF1.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1
AUTHORS Caldwell,R.B., Kierzek,A.M., Arakawa,H., Bezzubov,Y., Zaim,J.,
Fiedler,P., Kutter,S., Blagodataski,A., Kostovska,D., Koter,M.,
Plachy,J., Carnini,P., Hayashizaki,Y. and Buerstedde,J.M.,
Full-length cDNAs from chicken bursal lymphocytes to facilitate
gene function analysis
Genome Biol. 6 (1), R6 (2005)
15642098
REFERENCE 2 (bases 1 to 4339)
AUTHORS Caldwell,R.B.
Direct Submission
Submitted (20-MAY-2004) Caldwell R.B., GSF - Forschungszentrum,
Institut fuer Molekulare Strahlenbiologie, Ingolstaedter Landstr.
1, D-85764 Neuherberg, GERMANY
FEATURES
Location/Qualifiers
1..4339
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="CB"
/db_xref="taxon:9031"
/clone="6f12"
/cell_type="lymphocyte"
/tissue_type="bursa"
/clone_lib="rikeni"
/dev_stage="2 weeks old"
1..4339
/locus_tag="RCJMB04_6f12"
<1..261
/locus_tag="RCJMB04_6f12"
262..2844
/locus_tag="RCJMB04_6f12"
/notes="ORF1"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC31429.1"
/db_xref="GI:53129949"
/translocation="MEISSAPVGAQPLLMPRRPGVGTMGKPIKLANCFQVEIPKI
DVTLYEVDIKPKCKPRVNRVVDVSMVQHPKVTIFGDRPVYDGRKSLYANPLPVAT
TGVDDLTPLGEGKDRPKVSIKFKVRSVSHLLHLVLTGRTLTPELPPKPKPSTNPV
HAVDVVIRHLPSNMKXTPVGRSFPFSAPEYDHPILGGGVWFQGHQSVRPAWKKMLNI

```





TITLE Novel human genes and gene expression products ii  
JOURNAL Patent: JP 2002519000-A 2528 02-JUL-2002;  
COMMENT CHIRON CORP,RISEQ INC  
OS Homo sapiens (human)  
PN JP 2002519000-A/2528  
PD 02-JUL-2002  
PF 28-JAN-1999 JP 2000556580  
PR 28-JAN-1998 US 60/072910,24-FEB-1998 US 60/075954 PR.  
31-MAR-1998 US 60/080114,03-APR-1998 US 60/080515 PR  
03-APR-1998 US 60/080666,21-OCT-1998 US 60/105234 PR  
28-OCT-1998 US 60/105877  
PI LOUIS T WILLIAMS,JAIME ESCOBEDO,MICHAEL A INNIS,PABLO PI  
DOMINGUEZ GARCIA.  
PI JULIE SUDDUTH KLINGER,CHRISTOPH REINHARD,KLAUSE GIESE,PILIPPO  
PI RANDAZZO,  
PI GIULIA C KENNEDY,DAVID POT,ALTAP KASSAM,GEORGE LAMSON,RADOJE  
PI DRMANAC,  
PI RADOMIR CRKVENJAKOV,MARK DICKSON,SNEZANA DRMANAC,IVAN LABAT,  
PI DENA LESHKOWITZ,DAVID KITA,VERONICA GARCIA,LEE WILLIAM JONES,  
PI BIRJIT STACHE CRAIN  
PC C12N15/09,C12N15/09,C07K14/47,C07K14/82,C07K16/18,C12N1/15, PC  
C12N1/19,  
PC C12N1/21,C12N5/10,C12Q1/68,C12N15/00,C12N5/00,C12N15/00 CC n  
FH Key = A,T,C or G  
FT misc feature Location/Qualifiers  
source 1. .734  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

Query Match 1.9%; Score 62.4; DB 2; Length 734;  
Best Local Similarity 51.8%; Pred. No. 3.7e-06;  
Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 2698 GAATTGTGAACCCATCATCGGAACTCGGATAAACTTATCGTTTCGAAATACAAAT 2757  
|||||  
Db 209 GAATGGCAATATCCAGCTGGAAACAAGTTGATACAGACATTACACCCCATATGAT 268  
2758 TCGATTCTTCTCTGCTAGATGTGCGAAACCCATCTCGTTGCCCTTCCGGTTTCATTATG 2937  
|||||  
Db 269 TCGATTCTTCTCTGCTAGATGTGCGAAACCCATCTCGTTGCCCTTCCGGTTTCATTATG 448  
2818 CTGTTATGTATGACGATAAAGGATGAGCCAAAGATGAAGTCTATAAATGACCTACGGAC 2877  
|||||  
Db 329 ATGTTTATGGATGATACATCTGTTTACTGCAGATGAACCTTCAGCTGCTAACTTACCAGC 388  
2878 TTGCTTTCTCTGCTAGATGTGCGAAACCCATCTCGTTGCCCTTCCGGTTTCATTATG 2937  
|||||  
Db 389 TCTGCCACACTTACGTACGCTGTACACGATCTGTTCTATACCTGCACCGCGTATTATG 448  
2938 CTCAATTATCATGTGAAAGGCGAAAGAGCTT 2969  
|||||  
Db 449 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 480

RESULT 13  
AR767980 734 bp DNA linear PAT 08-DEC-2005  
LOCUS Sequence 2528 from patent US 6964868.  
DEFINITION AR767980  
ACCESSION AR767980.1 GI:83343259  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
Unclassified.  
1 (bases 1 to 734)  
Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,  
Sudduth-Klinger,J., Reinhard,C., Giese,K., Randazzo,F.,  
Kennedy,G.C., Pot,D., Kassam,A., Lamson,G., Drmanac,R.,  
Crkvenjakov,R., Dickson,M., Drmanac,S., Labat,I., Leshkowitz,D.,

TITLE Kita,D., Garcia,V., Jones,L.W. and Stache-Crain,B.  
JOURNAL Human genes and gene expression products II  
COMMENT Patent: US 6964868-A 2528 15-NOV-2005;  
Nuvelo, Inc.; Sunnyvale, CA  
FEATURES source 1. .734  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 1.9%; Score 62.4; DB 2; Length 734;  
Best Local Similarity 51.8%; Pred. No. 3.7e-06;  
Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 2698 GAATTGTGAACCCATCATCGGAACTCGGATAAACTTATCGTTTCGAAATACAAAT 2757  
|||||  
Db 209 GAATGGCAATATCCAGCTGGAAACAAGTTGATACAGACATTACACCCCATATGAT 268  
2758 TCGATTCTTCTGCTAGATGTGCGAAACCCATCTCGTTGCCCTTCCGGTTTCATTATG 2817  
|||||  
Db 269 TCGATTCTTCTGCTAGATGTGCGAAACCCATCTCGTTGCCCTTCCGGTTTCATTATG 328  
2818 CTGTTATGTATGACGATAAAGGATGAGCCAAAGATGAAGTCTATAAATGACCTACGGAC 2877  
|||||  
Db 329 ATGTTTATGGATGATACATCTGTTTACTGCAGATGAACCTTCAGCTGCTAACTTACCAGC 388  
2878 TTGCTTTCTCTGCTAGATGTGCGAAACCCATCTCGTTGCCCTTCCGGTTTCATTATG 2937  
|||||  
Db 389 TCTGCCACACTTACGTACGCTGTACACGATCTGTTCTATACCTGCACCGCGTATTATG 448  
2938 CTCAATTATCATGTGAAAGGCGAAAGAGCTT 2969  
|||||  
Db 449 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 480

RESULT 14  
BC025769 3040 bp mRNA linear PRI 20-OCT-2004  
LOCUS Homo sapiens eukaryotic translation initiation factor 2C, 3, mRNA  
DEFINITION (CDNA clone IMAGE:5210832).  
ACCESSION BC025769  
VERSION BC025769.1 GI:22213082  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 3040)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,  
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalon,D.K., Munz,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalios,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.D. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 3040)  
Director MGC Project.



**TITLE**  
**JOURNAL**  
 Direct Submission  
 Submitted (06-MAR-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
**REMARK**  
**COMMENT**  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: http://www.nisc.nih.gov/  
 Contact: nisc.mgc@nih.gov  
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaapi, R.,  
 Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
 Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAC Plate: 49 Row: j Column: 13  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 29294646.

**FEATURES**  
 source

Location/Qualifiers  
 1..3040  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5210832"  
 /tissue\_type="Blood, adult leukocytes"  
 /clone\_lib="NIH\_MGC\_118"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

**ORIGIN**

Query Match 1.9%; Score 62.4; DB 5; Length 3040;  
 Best Local Similarity 51.8%; Pred. No. 3.5e-06;  
 Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 2698 GAATTGTGAACCCATCATCCGGAACAACTGTGGATAAACTTATCGTTTCGAAATACAAAT 2757  
 DB 2091 GAAGTGGCAATATCCAGCTGGAAACAACTGTGATACAGACATTACACACCCATATGAGT 2150  
 QY 2758 TCGATTTTCTTGGCATCTCATCTGTTGTCCTTGGTACATCTCGTCCAGGACATTACA 2817  
 DB 2151 TCGATTTTACCTCTGTAGCCATGCTGGAATACAGGATACAGTCTGCTTCCACTATC 2210  
 QY 2818 CTGTTATGATGACGATAAGGAATCAGCAAGATCAAGTCTATAAATGACCTACGGAC 2877  
 DB 2211 ATGTTTTATGGGATGATACTGCTTTTCTGACAGATGAACTTACGCTGCTAACTTACCAG 2270  
 QY 2878 TTGCTTTTCTCTCTGCTAGATGTCGAAACCCATCTCGTTGGCTGTTCGGGTTCAATTATG 2937  
 DB 2271 TCTGCCACACTTACGTACGCTGTACAGATCTGTTCTATACCTTCGACCCAGCGTATTATG 2330  
 QY 2938 CTCATTATCATGTGAAAAGCGAAAGAGCTT 2969  
 DB 2331 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 2362

**RESULT 15**

BD158050  
 LOCUS 3050 bp DNA linear PAT 17-JAN-2003  
 DEFINITION  
 Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION  
 BD158050  
 VERSION  
 BD158050.1 GI:27863808

**KEYWORDS**  
 JP 2002191363-A/12893.  
**SOURCE**  
 Homo sapiens (human)  
**ORGANISM**  
 Homo sapiens  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
**REFERENCE**  
 1 (bases 1 to 3050)  
**AUTHORS**  
 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
**TITLE**  
 Primer for synthesizing full-length cDNA and use thereof  
**JOURNAL**  
 Patent: JP 2002191363-A 12893 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
**COMMENT**  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/12893  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
 10,  
 PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
**FT** CDS Location/Qualifiers  
 (101)..(2680).  
**FEATURES**  
 source  
 1..3050  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
**ORIGIN**  
 Query Match 1.9%; Score 62.4; DB 2; Length 3050;  
 Best Local Similarity 51.8%; Pred. No. 3.5e-06;  
 Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 2698 GAATTGTGAACCCATCATCCGGAACAACTGTGGATAAACTTATCGTTTCGAAATACAAAT 2757  
 DB 2280 GAAGTGGCAATATCCAGCTGGAAACAACTGTGATACAGACATTACACACCCATATGAGT 2339  
 QY 2758 TCGATTTTCTTGGCATCTCATCTGTTGTCCTTGGTACATCTCGTCCAGGACATTACA 2817  
 DB 2340 TCGATTTTACCTCTGTAGCCATGCTGGAATACAGGATACAGTCTGCTTCCACTATC 2399  
 QY 2818 CTGTTATGATGACGATAAGGAATGAGCAAGATGAGCTCTATAAATGACCTACGGAC 2877  
 DB 2400 ATGTTTTATGGGATGATAAATCTGCTTTTCTGACAGATGAACTTCAGCTGCTTAACCTAC 2459  
 QY 2878 TTGCTTTTCTCTCTGCTAGATGTCGAAACCCATCTCGTTGGCTGTTCGGGTTCAATTATG 2937  
 DB 2460 TCTGCCACACTTACGTACGCTGTACAGATCTGTTCTTATACCTTCGACCCAGCGTATTATG 2519  
 QY 2938 CTCATTATCATGTGAAAAGCGAAAGAGCTT 2969  
 DB 2520 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 2551

Search completed: July 6, 2006, 02:44:28  
 Job time : 17500 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2006, 21:40:42 ; Search time 577 Seconds  
(without alignments)  
10464.589 Million cell updates/sec

Title: US-10-645-746-2  
Perfect score: 3227  
Sequence: 1 cagccacaaagtcatgaac.....tttaaaaaaaaaaaaaa 3227

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: , 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/ina/1 COMB.seq:\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/ina/5 COMB.seq:\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/ina/6A COMB.seq:\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/ina/6B COMB.seq:\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/ina/7 COMB.seq:\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/ina/H COMB.seq:\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/ina/PCTUS COMB.seq:\*
- 8: /EMC\_Celerra\_SID33/ptodata/2/ina/PP COMB.seq:\*
- 9: /EMC\_Celerra\_SID33/ptodata/2/ina/RE COMB.seq:\*
- 10: /EMC\_Celerra\_SID33/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62.4	1.9	734	4	US-09-297-648-2528 Sequence 2528, Ap
2	62	1.9	7218	2	US-08-232-463-14 Sequence 14, Appl
3	61.4	1.9	771	3	US-09-248-796A-6819 Sequence 6819, Ap
4	54.2	1.7	1141	3	US-09-806-708B-22 Sequence 22, Appl
5	49.6	1.5	1141	3	US-09-806-708B-22 Sequence 22, Appl
6	47	1.5	2910	3	US-09-533-029-105 Sequence 105, App
7	45.6	1.4	2328	3	US-10-043-774B-1 Sequence 1, Appli
8	45.6	1.4	3472	3	US-09-873-737A-5 Sequence 5, Appli
9	45.2	1.4	300	4	US-09-297-648-1917 Sequence 1917, Ap
10	44.2	1.4	7478	3	US-09-949-016-948 Sequence 948, App
11	44.2	1.4	7478	3	US-09-949-016-4453 Sequence 4453, Ap
12	43	1.3	832	3	US-09-621-976-2813 Sequence 2813, Ap
13	42.8	1.3	723	3	US-09-270-767-4630 Sequence 4630, Ap
14	42.8	1.3	723	3	US-09-270-767-19912 Sequence 19912, A
15	42.8	1.3	5838	3	US-08-956-1718-66 Sequence 66, Appl
16	42.8	1.3	5838	3	US-08-781-986A-66 Sequence 66, Appl
17	42.6	1.3	601	3	US-09-949-016-197722 Sequence 197722, A
18	42.6	1.3	601	3	US-09-949-016-197723 Sequence 197723, A
19	42.6	1.3	85665	3	US-09-949-016-17345 Sequence 17345, A
20	42	1.3	19124	2	US-08-487-826B-13 Sequence 13, Appl
21	41.6	1.3	1875	7	PCT-US96-05320A-997 Sequence 997, App
22	41.6	1.3	1830121	3	US-09-557-884-1 Sequence 1, Appli
23	41.6	1.3	1830121	3	US-09-643-990A-1 Sequence 1, Appli

C 24	41.6	1.3	1830121	3	US-10-158-865-1	Sequence 1, Appli
C 25	41.4	1.3	11485	3	US-09-410-464-9	Sequence 9, Appli
C 26	41.4	1.3	114426	3	US-09-949-016-15078	Sequence 15078, A
C 27	41.4	1.3	163317	4	US-09-531-120-212	Sequence 212, App
C 28	41.2	1.3	601	3	US-09-949-016-196507	Sequence 196507, A
C 29	41.2	1.3	451924	3	US-09-949-016-12896	Sequence 12896, A
C 30	41.2	1.3	451925	3	US-09-949-016-17305	Sequence 17305, A
C 31	40.8	1.3	1809	4	US-09-880-107-3376	Sequence 3376, Ap
C 32	40.6	1.3	247299	3	US-09-949-016-17590	Sequence 17590, A
C 33	40.4	1.3	2820	7	PCT-US93-11725-1	Sequence 1, Appli
C 34	40.4	1.3	47375	3	US-09-949-016-15420	Sequence 15420, A
C 35	40.4	1.3	168575	3	US-09-426-290-1	Sequence 1, Appli
C 36	39.6	1.2	1664976	3	US-08-316-421B-1	Sequence 1, Appli
C 37	39.6	1.2	1664976	3	US-09-692-570-1	Sequence 119957, A
C 38	39.4	1.2	601	3	US-09-949-016-119957	Sequence 8910, Ap
C 39	39.4	1.2	1680	3	US-09-248-796A-8910	Sequence 11950, A
C 40	39.4	1.2	121427	3	US-09-949-016-11950	Sequence 13230, A
C 41	39.4	1.2	121433	3	US-09-949-016-13230	Sequence 13230, A
C 42	39.4	1.2	194889	3	US-09-949-016-15654	Sequence 15654, A
C 43	39.2	1.2	832	3	US-09-621-976-2813	Sequence 2813, Ap
C 44	39	1.2	381	3	US-09-270-767-408	Sequence 408, App
C 45	39	1.2	381	3	US-09-270-767-15690	Sequence 15690, A

#### ALIGNMENTS

##### RESULT 1

US-09-297-648-2528  
; Sequence 2528, Application US/09297648  
; Patent No. 6964868  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Escobedo, Jaime  
; APPLICANT: Innis, Michael A.  
; APPLICANT: Garcia, Pablo Dominquez  
; APPLICANT: Sudduth-Klinger, Julie  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Giese, Klaus  
; APPLICANT: Randazzo, Filippo  
; APPLICANT: Kennedy, Giulia C.  
; APPLICANT: Pot, David  
; APPLICANT: Kassan, Altaf  
; APPLICANT: Lamson, George  
; APPLICANT: Drmanac, Radoje  
; APPLICANT: Crkvenjakov, Radomir  
; APPLICANT: Dickson, Mark  
; APPLICANT: Drmanac, Snezana  
; APPLICANT: Labat, Ivan  
; APPLICANT: Leshkowitz, Dena  
; APPLICANT: Kita, David  
; APPLICANT: Garcia, Veronica  
; APPLICANT: Jones, William Lee  
; APPLICANT: Stache-Crain, Birgit  
; TITLE OF INVENTION: No. 6364868el Human Genes and Gene Expression  
; FILE REFERENCE: 2300-1481  
; CURRENT APPLICATION NUMBER: US/09/297,648  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR FILING DATE: 1998-01-28  
; PRIOR FILING DATE: 1998-01-28  
; PRIOR FILING DATE: 1998-01-28  
; PRIOR FILING DATE: 1998-02-24  
; PRIOR FILING DATE: 1998-02-24  
; PRIOR FILING DATE: 1998-04-03  
; PRIOR FILING DATE: 1998-04-03  
; PRIOR FILING DATE: 1998-04-03  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR FILING DATE: 1998-10-21  
; NUMBER OF SEQ ID NOS: 5252  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2528  
; LENGTH: 734  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1).....(734)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-297-648-2528

Query Match 1.9%; Score 62.4; DB 4; Length 734;  
Best Local Similarity 51.8%; Pred. No. 7.5e-06;  
Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
QY 2698 GAATGTGAACCATCATCGGACAACTGTGATAAATCTATCGTTTGAATACAAAT 2757  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
209 GAAGTGGCAATATCCAGCTGGAAACACAGTGTGATACAGACATTACACCCCATATGAGT 268  
QY 2758 TCGATTTTCTTGGCATCTCATCATGTGTCTTGGTACATCTCGTCGAGGACATTACA 2817  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
269 TCGATTTTACCTCTGTAGCATCTGGAATACAGGTACAGTCTGCTTACACTATC 328  
QY 2818 CTGTATGATGACGATAAGGAATGAGCAAGATGAAGTCTATAAANTGACCTACGGAC 2877  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
329 ATGTTTATGGGATGATAACTGCTTTACTGCAGATGAACCTTCAGCTGCTAACTTACCAGC 388  
QY 2878 TTGCTTTTCTCTGTAGATGTCGAAACCCCATCTCGTTGCTGTTCGGTTCATTATG 2937  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
389 TCTGCCACACTTACGTACCTGTACAGATCTGTTTCTATACCTGCACGAGGTATTATG 448  
QY 2938 CTCATTATCATGTGAAAGCGAAAGAGCTT 2969  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
449 CTCACCTGGTAGCATTTAGACCGCATATCAT 480

RESULT 2  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109

; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpc-Fls  
US-08-232-463-14  
Query Match 1.9%; Score 62; DB 2; Length 7218;  
Best Local Similarity 8.6%; Pred. No. 2.6e-05;  
Matches 38; Conservative 223; Mismatches 183; Indels 0; Gaps 0;  
QY 2240 TCCAGGTGGAATCTATCTATCGAAATATGATTGTGACTCAAGAAGATGTCGTCGCCGTGA 2299  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1437 TACRR 1378  
QY 2300 CGGTGCAGTGGCTCATGACGGGAAGACACATATTTTGGGAAGCAAGTTCGTGAAATT 2359  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1377 RRR 1318  
QY 2360 GCTCAGAGAATTCGAGAAACACACGACAATCGAGCACCGCATATTTGTAGTCTTATCG 2419  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1317 RRR 1258  
QY 2420 ACAGCGATTAGCGATTGCGAGATGCTAGTGTAGTCTAGTATGATGAGCTTCGATCTTAA 2479  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1257 RRR 1198  
QY 2480 AAGCGAAGTAAACAATTCATCTCGAACGGGATGGAGAAGATCCAGAGCCGAAGTACAC 2539  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1197 RRR 1138  
QY 2540 GTTCATTGTGATTCAGAAAGACACAATACAGATTCGTTTCGAAAGATCGAAAGATAA 2599  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1137 RRR 1078  
QY 2600 GCCAGTGTCAATAAAGATCTTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2659  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1077 RRRRRRRRRRRATCGAAGCTCCCTCGACCTGACGCAAGCTCGGAATTAATTCGTGTGAG 1018  
QY 2660 ACAATGGGAGGAGGATATGAAAGA 2683  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1017 CGTATGGCAACGAAAGGAAAATA 994

RESULT 3  
US-09-248-796A-6819  
; Sequence 6819, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 6819  
; LENGTH: 771  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-6819

Query Match 1.9%; Score 61.4; DB 3; Length 771;  
Best Local Similarity 50.9%; Pred. No. 1.4e-05;  
Matches 146; Conservative 0; Mismatches 141; Indels 0; Gaps 0;



```
QY 192 TACTACGAGTATGAAGTGAATGCAAAAGGAGTATTGAATAGAAAAACCCAGAAACCT 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 849 WMTRTNNNNNNNNNCRNTRTWABWBSWCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 252 TTCCCAAAAAGACAGAAATTCCTCAATTCGCGTTCGCAAACTCTCTCGCAACATCTT 311
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 789 RANNNNAARMARTCNYYHAAVTTHTDWCYKTNWNTWYDMTMBTTTTNNMTTSN 730
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 312 CGCGATGAGAAGACAGACAGATTTTCTCGAAGACTATGTTTTGATGAAAGGAC 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 729 MTNNNNNNWACTNNNNNNKAYAHATNNNGCMNNTDARTNNTTVMRNRMTNKT 670
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 ACTGTTATAGTTGTGCTGACTGAACACTGTCACATCAAAATGCTGTTTCGGAGAA 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 669 RYSTTRRHGTGATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 432 GTAGTAAAAAGGATTCGGAGAAAAAGATGAAAGGATTCGAGAAAAAATCTTATAC 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 609 DTTCTVDVWASWVWYANWRCRDVTYTRNTYKSYAHSWYWSNMMYRRYSARN 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 492 ACAATGACTTACCTATCGTAAAAAATTTCACTGAACTTTAGTCGAGAAAAATCCGAA 551
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 WSSMARWTTNNNNWMSGBVBRWAGTMMWRHNNNNNTDTRYWWRWARBTITTVYDSMC 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 552 AAAGACGAAGACGGAATCGGAGTTACAAATTCCTCGAAGATGTTATGACCCAGAAAT 611
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 NAKSMWRGNWRAWMKMMANNDAGAMDHTYWMGNNTMMRRMMWMMNMAWCRAYCCN 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 612 CGTACGCGCTTTTCTGACAGAGAGATTAAGATACAATTCGCGAAAAATTTTGTGTAC 671
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 NNNNRACVHKHWRWTKYMKKAAACNNNBKAMMYRVAMMYSEDTNTTDMWMTSDW 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 672 GATAATAATTCATCTCGAGTTCCTGAACTCCTGAACTGTTACAGATCCAAACAGATTCGAACAA 731
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 BWHVTVDTVMRAWNNNNNNNRBCKTTSWMMMDHNNTHCTYGNNTWGSAYBNAAISM 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 732 TCATTAGAAGTAGCACAAGATCGAAGCATGTTTGGAAATTACATTCGAATCAAGAA 791
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 WAAGASBNVTYNNCWRMTYMGKMTNNNNNNKAWYR-TKTVAWCNNRYYYDTAVWTKR 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 792 TTGTTGAGTGGTGAACCTGTGCTCAATTTTGAATTTGCAATTTGCGATAAACTATTCTACAATGCA 851
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 NYKYCYAYBYBYMYGKHBNWRBHRBSNMMWVKCNKYMYSHVHAMRYKBABA 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 852 CGGAAATCTCTCTCTGAGTATCTTCTCCTAAATTTGTCGACCCCGAGTCGTGTAACGAT 911
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 VGCNNWKMMAHHHWCATNNNNMMWYAYMHMHKKGAATNNKNTABRDDHBAHVKT 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 912 GATGACGAAAGAGTCTTAAACAAACTGATCGCGGGAAGAAATGACATCAGACAAGCC 971
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 WYWRDYWCAMCMWNAKAVRTAMKHMYTDRYVSANNTGVRWMMWMMWYSHNNRW 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 972 GCGCGGCCAAGAAATCG 988
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 YRMGRKYTWAMYSMS 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 6
US-09-533-029-105
; Sequence 105, Application US/09533029
; Patent No. 666446
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddie, James
; APPLICANT: Pinede, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
```

```
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 105
; LENGTH: 2910
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1149
US-09-533-029-105

Query Match 1.5%; Score 47; DB 3; Length 2910;
Best Local Similarity 50.2%; Pred. No. 0.12;
Matches 116; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 2714 ATCCGAAACAACTGTGGATAAACTTATCGTTTCGAAATACAAATTCGATTTTCTTGGC 2773
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2529 ACCAGTACTGTCGTGGACACTAAATCTGTCAACCTAATGAATTCGACTTCTATTGAA 2588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2774 ATCTCATCATGCTGCTTGTGTACATCTCGTCCAGGACATTACACTGTTATGTATGACGA 2833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2589 CAGCCATGCTGTAATTCAGGGAAACAAGCAGCGCGGCACATTACCATGTACTTCTCGATGA 2648
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2834 TAAAGGATGAGCCAGATGAAGTCTATAAATGACCTACGAGCTGCTTTTCTCTCTCG 2893
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2649 GAACGGTTTACCGCTGATCAGTGCATAATGCTCAAAACAACCTCTGCTACAGTATGC 2708
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2894 TAGATGTCGAAACCCATCTCGTTGCCCTGTTCCCGTTTCATTATGCTCATTT 2944
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2709 GAGGTGTACAAATCTGTCTCAATTTGTCACCAGCCTACTAGCTCACTT 2759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-10-043-774B-1
; Sequence 1, Application US/10043774B
; Patent No. 6900017
; GENERAL INFORMATION:
; APPLICANT: University of Illinois at Chicago
; APPLICANT: Sharma, Arun
; APPLICANT: Hoffman, Ronald
; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES
; FILE REFERENCE: MBHB: CU08/PPA
; CURRENT APPLICATION NUMBER: US/10/043,774B
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2328)
; OTHER INFORMATION: Human Hiwi Protein
US-10-043-774B-1

Query Match 1.4%; Score 45.6; DB 3; Length 2328;
Best Local Similarity 49.2%; Pred. No. 0.25;
Matches 120; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 2698 GAATTTGTAACCCATCATCCGGAACAACCTGTGATAAACTTATCGTTTCGAAATACAAAT 2757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2006 GACTTCAGAAATCCATCTTCTGGAACAGTTATTGATGTAGAGTTACCAGACCAAGATGGT 2065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2758 TCGATTTTCTTGGCATCTCATGCTGCTGCTTGGTGTCTGTACATCTCGTCCAGGACATTACA 2817
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```











GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2006, 21:15:33 ; Search time 1721 Seconds  
(without alignments)  
13073.472 Million cell updates/sec

Title: US-10-645-746-2  
Perfect score: 3227  
Sequence: 1 cagccacaagtgtatgaac.....tttaaaaaaaaaaaaaaa 3227

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_8:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*  
14: Geneseqn2005s:\*  
15: Geneseqn2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3227	100.0	3227	4	AAS03283 C. elegans
2	2497.4	77.4	3709	4	AAS03282 C. elegans
3	63.8	2.0	2827	9	ADB17459 Soybean p
4	63.8	2.0	2827	14	Aec75683 Soybean Z
5	62.4	1.9	490	9	ACH23411 Human adu
6	62.4	1.9	734	2	AAZ15059 Human gen
7	62.4	1.9	2568	12	ADJ94706 Human euk
8	62.4	1.9	2914	4	AAS3425 Human pol
9	62.4	1.9	2914	8	ABX13625 Human cyt
10	62.4	1.9	3050	4	AAH16058 Human cdn
11	62.4	1.9	3050	12	ADH77114 Human PAZ
12	62.4	1.9	3325	8	AAD55860 Human nuc
13	62.4	1.9	3736	8	ABX34752 Human mdd
14	61.8	1.9	2679	13	ADQ80688 Arabidops
15	60	1.9	1501	9	ADB17461 Soybean p
16	60	1.9	1501	14	AEC75685 Soybean Z
17	57	1.8	441	9	ACH28746 Human adu
18	56.4	1.7	2772	12	ADJ94705 Human euk

19	54.8	1.7	2951	4	AAH24562	Aah24562 Translati
20	54	1.7	2000	11	ACL35363	ACL35363 Rice stre
21	53	1.6	3520	13	ADQ80702	Adq80702 Arabidops
22	52.8	1.6	3096	9	ADB17463	Adb17463 Corn post
23	52.8	1.6	3096	14	AEC75687	Aec75687 Corn Argo
24	51.4	1.6	1565	9	ADB17505	Adb17505 Wheat pos
25	51.4	1.6	1565	14	AEC75729	Aec75729 Wheat Arg
26	51	1.6	1158	3	AAC50998	Aac50998 Arabidops
27	51	1.6	3120	13	ADQ80694	Adq80694 Arabidops
28	51	1.6	3147	12	ADI43622	Adi43622 Plant tra
29	51	1.6	3195	3	AAC50992	Aac50992 Arabidops
30	51	1.6	3326	3	AAC47079	Aac47079 Arabidops
31	50.2	1.6	1238	9	ADB17487	Adb17487 Rice post
32	50.2	1.6	1238	14	AEC75711	Aec75711 Rice Argo
33	49.8	1.5	2960	13	ADQ80704	Adq80704 Arabidops
34	49.8	1.5	2973	12	ADO61720	Ado61720 Transcrip
35	49.8	1.5	3008	12	ADO63642	Ado63642 Transcrip
36	49.6	1.5	2622	13	ADQ80690	Adq80690 Arabidops
37	48.8	1.5	2775	13	ADQ80686	Adq80686 Arabidops
38	48.8	1.5	2793	6	ABZ12589	Abz12589 Arabidops
39	48.8	1.5	2793	10	ADE31434	Ade31434 Plant vie
40	48.8	1.5	2793	12	ADI44098	Adi44098 Plant tra
41	48.6	1.5	2000	8	ADA71938	Ada71938 Rice gene
42	47.8	1.5	2960	13	ADQ80696	Adq80696 Arabidops
43	47.8	1.5	2967	3	AAC50603	Aac50603 Arabidops
44	47.8	1.5	2967	6	ABZ12627	Abz12627 Arabidops
45	47.8	1.5	3071	14	ADY65087	Ady65087 S. manson

#### ALIGNMENTS

#### RESULT 1

AAS03283

ID AAS03283 standard; cDNA; 3227 BP.

XX AAS03283;

AC AAS03283;

DT 07-SEP-2001 (first entry)

DE C. elegans cDNA encoding RNA interference pathway protein RDE-1.

XX RNA interference; RNAi; RDE-1; genetic interference; antibody; dsRNA;

KW double-stranded RNA-dependent gene silencing; ss.

XX Caenorhabditis elegans.

OS Caenorhabditis elegans.

XX Key Location/Qualifiers

FT 5'UTR 1..20

FT /\*tag= a

FT CDS 21..3083

FT /\*tag= b

FT /product= "RDE-1"

XX

PN WO200129058-A1.

XX

PD 26-APR-2001.

XX

PF 13-OCT-2000; 2000WO-US028470.

XX

PR 15-OCT-1999; 99US-0159776P.

XX

PR 30-MAR-2000; 2000US-0193218P.

XX

PA (UWMA-) UNIV MASSACHUSETTS.

XX (CARN-) CARNEGIE INST WASHINGTON.

PI Mello CC, Fire A, Tabara H, Grishok A;

XX WPI; 2001-316239/33.

DR P-PSDB; AAU01856.

XX Novel RNA interference pathway genes and their protein products involved

PT in mediation of genetic interference, useful for modulating and studying

PT regulation of RNA interference pathway.

XX Claim 1; Fig 6; 76pp; English.

XX The sequence encodes the RNA interference (RNAi) pathway protein RDE-1. RDE-1 and RDE-4 are involved in the pathway mediating double-stranded RNA-dependent gene silencing (genetic interference) RDE-1 and RDE-4 protein is useful for preparing an RNAi agent, by incubating a dsRNA in the presence of the proteins. The prepared RNAi agents can be used as sequence-specific interfering agents for targeted genetic interferences. The nucleic acids are useful for studying the regulation of RNAi pathway and to generate knockout strains of animals such as C.elegans. RDE-1 and RDE-4 genes and their products are useful for modulating RNAi pathway activity. The polypeptides are useful for generating and testing antibodies specific for the polypeptides which are useful for studying the RNAi pathway in C.elegans and other organisms. RNAi pathway genes are useful for mediating specific processes, e.g. a gene that mediates dsRNA uptake by the cells is useful for transporting other RNAs into cells or for facilitating entry of agents such as drugs into cells. RNAi pathway mutant strains (rde-1, rde-4 mutants) are useful in genetic screens to identify additional RNAi pathway components

XX Sequence 3227 BP; 1145 A; 575 C; 666 G; 841 T; 0 U; 0 Other;

Query Match 100.0%; Score 3227; DB 4; Length 3227;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CAGCCACAAAGTGATGAACATGCTCCTCGAAATTTTCCGAAATTTGGAAAAAGGATTTTATC	60
DB	1	CAGCCACAAAGTGATGAACATGCTCCTCGAAATTTTCCGAAATTTGGAAAAAGGATTTTATC	60
QY	61	GTCAATTCCTCGATCGGAGATGAATGGCTTCGAGGGCCCACTGTGTAATCGGACGGCA	120
DB	61	GTCAATTCCTCGATCGGAGATGAATGGCTTCGAGGGCCCACTGTGTAATCGGACGGCA	120
QY	121	AATTCCTATGAGAGAAGTACTCTTTTGTGTAATTTGGTTCAAGTTCTCCAGCAAAATTT	180
DB	121	AATTCCTATGAGAGAAGTACTCTTTTGTGTAATTTGGTTCAAGTTCTCCAGCAAAATTT	180
QY	181	ACGATCGGGAATACTACGAGTATGAAGTGAATGAATGAAGGAAGTATTGAATAGAAAAAC	240
DB	181	ACGATCGGGAATACTACGAGTATGAAGTGAATGAATGAAGGAAGTATTGAATAGAAAAAC	240
QY	241	CAGGAAAACTTTCCCAAAAAAGACAGAAATTCCTCAATTCCTCGATTCGTCGCAAACTCTTCT	300
DB	241	CAGGAAAACTTTTCCCAAAAAAGACAGAAATTCCTCAATTCCTCGATTCGTCGCAAACTCTTCT	300
QY	301	GGCAACATCTTCGGCATGAGAGAGACAGACAGATTTTATCTCGAAGACTATGTTTTTG	360
DB	301	GGCAACATCTTCGGCATGAGAGAGACAGACAGATTTTATCTCGAAGACTATGTTTTTG	360
QY	361	ATGAAAGGACACTGTTTATAGTTTGTGCACTGAACACTGTGCATCAAAATGCTCG	420
DB	361	ATGAAAGGACACTGTTTATAGTTTGTGCACTGAACACTGTGCATCAAAATGCTCG	420
QY	421	TTTCGAGAAAGTAGTAAAGGATTCGAGAAAAAGATGAAGGAATTTGGAGAAAA	480
DB	421	TTTCGAGAAAGTAGTAAAGGATTCGAGAAAAAGATGAAGGAATTTGGAGAAAA	480
QY	481	AAATCTTATACAAATGATCTTACCTATCGTAAATAATTTCCACCTGAACTTTAGTCGAG	540
DB	481	AAATCTTATACAAATGATCTTACCTATCGTAAATAATTTCCACCTGAACTTTAGTCGAG	540
QY	541	AAAAATCGGAAAAAGACGAGCGAATCGGAGTTTACAAATTCCTGGAAGATGTTTATGA	600
DB	541	AAAAATCGGAAAAAGACGAGCGAATCGGAGTTTACAAATTCCTGGAAGATGTTTATGA	600
QY	601	CCGAGAAAGTTCGCTACGGCCCTTTTGTGAACGAGGAGATTTAAAGTACAAATTCGCGAAA	660
DB	601	CCGAGAAAGTTCGCTACGGCCCTTTTGTGAACGAGGAGATTTAAAGTACAAATTCGCGAAA	660
QY	661	ATTTGTGTACGATATAATTCATTTCTCGAGTTCTCTGAAATCGTTTTCAGATCCAAACA	720

Db 1741 TGTACGACGCGAAGAAATGAATATGCTTCTTACAAAATTTGTACATAAATACCGGAA 1800  
Qy 1801 TCGGTAGATTGAAATAGCCGCAACAGAGCGAAGAATAATGTTTGAACGCTTCCCGATA 1860  
Db 1801 TCGGTAGATTGAAATAGCCGCAACAGAGCGAAGAATAATGTTTGAACGCTTCCCGATA 1860  
Qy 1861 AAGAAACAAAAGTCTTAATGTTTCAATATCAATTTCCAAACGACAACTCAATGCTTACCGTT 1920  
Db 1861 AAGAAACAAAAGTCTTAATGTTTCAATATCAATTTCCAAACGACAACTCAATGCTTACCGTT 1920  
Qy 1921 TTGTGAACCAATTAATGCGATCACACCATCGGTGTAGCTAATTCAGCATATTAATTTCTGAAA 1980  
Db 1921 TTGTGAACCAATTAATGCGATCACACCATCGGTGTAGCTAATTCAGCATATTAATTTCTGAAA 1980  
Qy 1981 CAGTCACAAAAGCTTTGGCATCACTAAGGCAACGAAAGGATCAAAACGAATTTCTATC 2040  
Db 1981 CAGTCACAAAAGCTTTGGCATCACTAAGGCAACGAAAGGATCAAAACGAATTTCTATC 2040  
Qy 2041 AAATTGCAATGAAAATCAACGCGAAATTTAGAGGTATTTAACCCAGGAGCTTGACTGGTCAG 2100  
Db 2041 AAATTGCAATGAAAATCAACGCGAAATTTAGAGGTATTTAACCCAGGAGCTTGACTGGTCAG 2100  
Qy 2101 AAATTGCAAGAAATATCAACGAGAAAGAAAGAGACGGAAGCAATGCCATTAATCTATGT 2160  
Db 2101 AAATTGCAAGAAATATCAACGAGAAAGAAAGAGACGGAAGCAATGCCATTAATCTATGT 2160  
Qy 2161 ATGTTGGAATGATGTAATCACTCCAACTCTCTACAGTGAAATGATATTTATATAGCGG 2220  
Db 2161 ATGTTGGAATGATGTAATCACTCCAACTCTCTACAGTGAAATGATATTTATATAGCGG 2220  
Qy 2221 CTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGACTCAAG 2280  
Db 2221 CTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGACTCAAG 2280  
Qy 2281 AAGAAATGCTCCGCTGAGCGTGCAGTGGCTCATGGACGGGAAAGAAACAGATATTTGG 2340  
Db 2281 AAGAAATGCTCCGCTGAGCGTGCAGTGGCTCATGGACGGGAAAGAAACAGATATTTGG 2340  
Qy 2341 AAGCAAAAGTTCGTGAAATTCCTCAGAGAAATTCGAGAAATTCGAGAAATTCGAGCAAC 2400  
Db 2341 AAGCAAAAGTTCGTGAAATTCCTCAGAGAAATTCGAGAAATTCGAGCAAC 2400  
Qy 2401 CGCATATGAGTCTACGAGACGGAGTTAGCGATTCGGAGATGCTACGTTGTAGTCATG 2460  
Db 2401 CGCATATGAGTCTACGAGACGGAGTTAGCGATTCGGAGATGCTACGTTGTAGTCATG 2460  
Qy 2461 ATGAGCTTCGATCTTTAAAGGCGAAGTAAACCAATTCATGTCGGAACGGGATGGAGAAG 2520  
Db 2461 ATGAGCTTCGATCTTTAAAGGCGAAGTAAACCAATTCATGTCGGAACGGGATGGAGAAG 2520  
Qy 2521 ATCCAGAGCGGAAGTACAGCTTCATGTTGATTCAGAAAGACACAATACACGATTTGCTTC 2580  
Db 2521 ATCCAGAGCGGAAGTACAGCTTCATGTTGATTCAGAAAGACACAATACACGATTTGCTTC 2580  
Qy 2581 GAGAAATGGAAGAAGATAGCCAGTGGTCAATTAAGATCTTACTCTGCTGGAACAGATG 2640  
Db 2581 GAGAAATGGAAGAAGATAGCCAGTGGTCAATTAAGATCTTACTCTGCTGGAACAGATG 2640  
Qy 2641 TCGCTGTGCTGCTGTTTAAACCAATGGGAGGAGATATGAAGAAAGCAAAAGAACTGGAA 2700  
Db 2641 TCGCTGTGCTGCTGTTTAAACCAATGGGAGGAGATATGAAGAAAGCAAAAGAACTGGAA 2700  
Qy 2701 TTGTGAACCAATCATCCGGAACAACTGTGGATAACTTATCGTTTTCGAAATACAAATTCG 2760  
Db 2701 TTGTGAACCAATCATCCGGAACAACTGTGGATAACTTATCGTTTTCGAAATACAAATTCG 2760  
Qy 2761 ATTTTTTCTGGCATCTCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
Db 2761 ATTTTTTCTGGCATCTCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
Qy 2821 TTATGTATGACGATAAAGGAATGAGCAAGATGAAGTCTATAAATGACCTACGGACTTG 2880  
Db 2821 TTATGTATGACGATAAAGGAATGAGCAAGATGAAGTCTATAAATGACCTACGGACTTG 2880

Qy 2881 CTTTCTCTCTGCTAGATGTCGAAACCCATCTCGTTCCCTGTTCCGTTTCATTATGCTC 2940  
Db 2881 CTTTCTCTCTGCTAGATGTCGAAACCCATCTCGTTCCCTGTTCCGTTTCATTATGCTC 2940  
Qy 2941 ATTATCATCTGGAAGAGCGAAGAGAGCTTTATCGAACTTACAGGACATTTACATCGGTG 3000  
Db 2941 ATTATCATCTGGAAGAGCGAAGAGAGAGCTTTATCGAACTTACAGGACATTTACATCGGTG 3000  
Qy 3001 ACTATGACAGCCAGGACTCGACACGAAATGGAACATTTTCTCCAACTAAACGTGAAGT 3060  
Db 3001 ACTATGACAGCCAGGACTCGACACGAAATGGAACATTTTCTCCAACTAAACGTGAAGT 3060  
Qy 3061 ACCCTGGAATGCTGCTTCGCATACATTTTGCAAAAGTGTGCCCGTTTCAATCAAAATTTT 3120  
Db 3061 ACCCTGGAATGCTGCTTCGCATACATTTTGCAAAAGTGTGCCCGTTTCAATCAAAATTTT 3120  
Qy 3121 TCAATTGCTAGATATTTGCTTACTTTTAAAGCCGGTTTCAAAATTTCAATTTCCATG 3180  
Db 3121 TCAATTGCTAGATATTTGCTTACTTTTAAAGCCGGTTTCAAAATTTCAATTTCCATG 3180  
Qy 3181 ACTAACGTTTTCATAAATTAATTTGAAATTTTAAAAAATTTAAAAAATTTAAAAA 3227  
Db 3181 ACTAACGTTTTCATAAATTAATTTGAAATTTTAAAAAATTTAAAAAATTTAAAAA 3227

RESULT 2  
AAS03282

ID AAS03282 standard; DNA; 3709 BP.

XX AC AAS03282;

XX DT 07-SEP-2001 (first entry)

XX DE C. elegans genomic DNA encoding RNA interference pathway protein RDE-1.

XX KW RNA interference; RNAi; RDE-1; genetic interference; antibody; dsRNA;

XX OS double-stranded RNA-dependent gene silencing; ds.

XX CA Caenorhabditis elegans.

XX FH Key Location/Qualifiers

XX FT CDS 21..3582

XX FT /\*tag= a

XX FT /product= "RDE-1"

XX FT /note= "This sequence contains introns"

XX PN WO200129058-A1.

XX PD 26-APR-2001.

XX PF 13-OCT-2000; 2000WO-US028470.

XX PR 15-OCT-1999; 99US-0159776P.

XX PR 30-MAR-2000; 2000US-0193218P.

XX XX (UYMA-) UNIV MASSACHUSETTS.

XX PA (CARN-) CARNEGIE INST WASHINGTON.

XX XX Mello CC, Fire A, Tabara H, Grishok A;

XX PI WPI; 2001-316239/33.

XX DR P-PSDB; AAU01856.

XX XX Novel RNA interference pathway genes and their protein products involved

XX PT in mediation of genetic interference, useful for modulating and studying

XX PT regulation of RNA interference pathway.

XX XX Claim 1; Fig 5; 76pp; English.

XX CC The sequence is genomic DNA encoding the RNA interference (RNAi) pathway

XX CC protein RDE-1. RDE-1 and RDE-4 are involved in the pathway mediating

XX CC double-stranded RNA-dependent gene silencing (genetic interference) RDE-1

CC and RDE-4 protein is useful for preparing an RNAi agent, by incubating a  
 CC dsRNA in the presence of the proteins. The prepared RNAi agents can be  
 CC used as sequence-specific interfering agents for targeted genetic  
 CC interferences. The nucleic acids are useful for studying the regulation  
 CC of RNAi pathway and to generate knockout strains of animals such as  
 CC C.elegans. RDE-1 and RDE-4 genes and their products are useful for  
 CC modulating RNAi pathway activity. The polypeptides are useful for  
 CC generating and testing antibodies specific for the polypeptides which are  
 CC useful for studying the RNAi pathway in C.elegans and other organisms.  
 CC RNAi pathway genes are useful for mediating specific processes, e.g. a  
 CC gene that mediates dsRNA uptake by the cells is useful for transporting  
 CC other RNAs into cells or for facilitating entry of agents such as drugs  
 CC into cells. RNAi pathway mutant strains (rde-1, rde-4 mutants) are useful  
 CC in genetic screens to identify additional RNAi pathway components  
 XX  
 SQ Sequence 3709 BP; 1308 A; 621 C; 741 G; 1039 T; 0 U; 0 Other;

Query Match 77.4%; Score 2497.4; DB 4; Length 3709;  
 Best Local Similarity 85.8%; Pred. No. 0;  
 Matches 3188; Conservative 0; Mismatches 16; Indels 511; Gaps 17;  
 QY 1 CAGCCACAAAGTGATGAACATGCTCTCGAATTTTCCCGAATTGGAAAAAGGATTTTATC 60  
 DB 1 CAGCCACAAAGTGATGAACATGCTCTCGAATTTTCCCGAATTGGAAAAAGGATTTTATC 60  
 QY 61 GTCAATTTCTCGATCC----- 76  
 DB 61 GTCAATTTCTCGATCCGGTATGATCAATTTATGACAGCTATTAAGATATATAAGTTTGATA 120  
 QY 77 -----GGAGATGAATGGCTTGGAGGCCCACTGGTAAATCGACGCGCAAAATTC 125  
 DB 121 TTAATATTATAGGATGAATGGCTTGGAGGCCCACTGGTAAATCGACGCGCAAAATTC 180  
 QY 126 TATGAGAAGAGTACTTCTTTGGTAAATGGTTCGAAGTCTCCAGCAAAATTTACGAT 185  
 DB 181 TATGAGAAGAGTACTTCTTTGGTAAATGGTTCGAAGTCTCCAGCAAAATTTACGAT 240  
 QY 186 CGGGAATACACGATGATCAAGTGAATGACAAAGGAAGTATTGAATAGAAACACGGA 245  
 DB 241 CGGGAATACACGATGATCAAGTGAATGACAAAGGAAGTATTGAATAGAAACACGGA 300  
 QY 246 AAACCTTTCCCAAAAAGACAGA----- 268  
 DB 301 AAACCTTTCCCAAAAAGAGAAATTCCAATGTAAGTCTTGTAATTTAGTCAAAACTAAT 360  
 QY 269 ----AATCCAAATCCCGATCGTGCAAAACTCTTTGGCAACATCTTCGGCATGAGAGAA 325  
 DB 361 TTTATTTTTCAGTCCCGATCGTGCAAAACTCTTTGGCAACATCTTCGGCATGAGAGAA 420  
 QY 326 GCAGACAGATTTTATTTCTCGAGACTATGTTTTTGTGATGAAGGACACCTGTTTATAGTCT 385  
 DB 421 GCAGACAGATTTTATTTCTCGAGACTATGTTTTTGTGATGAAGGACACCTGTTTATAGTCT 480  
 QY 386 TTGTGACTGAACACTGTGCATCAAAAATGCTGGTTTCGGAGAAAGTAGTAAAAAGGA 445  
 DB 481 TTGTGACTGAACACTGTGCATCAAAAATGCTGGTTTCGGAGAAAGTAGTAAAAAGGA 540  
 QY 446 TTCGGAGAAAAAGATGAAGAGTTTGGAGAAAAAATCTTTATACAAATGATACCTTAC 505  
 DB 541 TTCGGAGAAAAAGATGAAGAGTTTGGAGAAAAAATCTTTATACAAATGATACCTTAC 600  
 QY 506 CTATCTGTAATAAATTTCACTGAACCTTTAGTCGAGAAATTCGGGAAAAAGACGAGAAGC 565  
 DB 601 CTATCTGTAATAAATTTCACTGAACCTTTAGTCGAGAAATTCGGGAAAAAGACGAGAAGC 660  
 QY 566 GAATCGGAGTTACAAATTTCTCTGA----- 588  
 DB 661 GAATCGGAGTTACAAATTTCTCTGAAGTTTATGAABAAACAGCATTTATAACAAACATTAG 720  
 QY 589 -----AGAATGTTATGACCCAGAAAGTTTCGCTACGGCCCTTTTGTGAACGAGAGATTAA 643  
 DB 721 CTTTTCAGAAATGTTATGACCCAGAAAGTTTCGCTACGGCCCTTTTGTGAACGAGAGATTAA 780

QY 644 ACT-----ACA 649  
 DB 781 AGTGTGAGTTGCAATTAATTAATTAATCACTCACTCATTTATATATTTTAAGACA 840  
 QY 650 ATTCCGAAAAATTTTGTGTACGATAAATAATTAATTTCTCGAGTTCTCGAATCGTTTCA 709  
 DB 841 ATTCCGAAAAATTTTGTGTACGATAAATAATTAATTTCTCGAGTTCTCGAATCGTTTCA 900  
 QY 710 CGATCCAAACAGATTGCAACATCATTAGAACTAGCAGCAACCAAGNATCGAAGCATGGTTGG 769  
 DB 901 CGATCCAAACAGATTGCAACATCATTAGAACTAGCAGCAACCAAGNATCGAAGCATGGTTGG 960  
 QY 770 AATTACATTTGGAATCAAAAGAAATTTGTCGATGGTGAACCTGTGCTCAATTTTGGCAA- 825  
 DB 961 AATTACATTTGGAATCAAAAGAAATTTGTCGATGGTGAACCTGTGCTCAATTTTGGCAAGTAA 1020  
 QY 826 -----TTGTCGATAAACTATTTC 842  
 DB 1021 GTTTGAGAAACCTGCGATAAAAAATCATGTGATTTTGTGTAAGTTTGTGATAAACTATTTC 1080  
 QY 843 TACAATGCAACCGAAATGCTCTCTCTGGATTTATCTTCTCTTAATTTGTCGACCCCGAGTCG 902  
 DB 1081 TACAATGCAACCGAAATGCTCTCTCTGGATTTATCTTCTCTTAATTTGTCGACCCCGAGTCG 1140  
 QY 903 TGTAAACGATGATGTACGAAAGATCTTAAACAAAAAATCATGTCGCGGAAAAAATGACAAATC 962  
 DB 1141 TGTAAACGATGATGTACGAAAGATCTTAAACAAAAAATCATGTCGCGGAAAAAATGACAAATC 1200  
 QY 963 AGACAAGCCGCGCGCCCAAGAAATTCGACAAATTTTGGAAAAATTTGAAGCTGAAATCGGCA 1022  
 DB 1201 AGACAAGCCGCGCGCCCAAGAAATTCGACAAATTTTGGAAAAATTTGAAGCTGAAATCGGCA 1260  
 QY 1023 GAAGTTTGGGATAACGAA----- 1040  
 DB 1261 GAAGTTTGGGATAACAAATTTTGTGTTTAAATTTATTCAAACAAATTAATATACAAATTTGATT 1320  
 QY 1041 ---ATGTGCGAGATTGACAGAACGACATCTGACATTTCTAGATTTTGTGCGAGGAAAACTCT 1097  
 DB 1321 TTCAGTTCGAGATTGACAGAACGACATCTGACATTTCTAGATTTTGTGCGAGGAAAACTCT 1380  
 QY 1098 CTTGTTTATAAAGTCACTGGTAAATCGGACAGAGAAAGAAATGCAAAAAAATGACGATACT 1157  
 DB 1381 CTTGTTTATAAAGTCACTGGTAAATCGGACAGAGAAAGAAATGCAAAAAAATGACGATACT 1440  
 QY 1158 ACATTTGTTCAAAATCTATGAGGAAACCAAAAGATTCATTGAGTTTCCCGACCTTACCACATA 1217  
 DB 1441 ACATTTGTTCAAAATCTATGAGGAAACCAAAAGATTCATTGAGTTTCCCGACCTTACCACATA 1500  
 QY 1218 GTCAAAGTTTAAAGTGGAGCAAAAAGAAATACGCTGTACCAATGGAAACATCTTGAAGTTCAAT 1277  
 DB 1501 GTCAAAGTTTAAAGTGGAGCAAAAAGAAATACGCTGTACCAATGGAAACATCTTGAAGTTCAAT 1560  
 QY 1278 GAGAGCCCAAAAAGATACAGAAATCGAATTTGATCTGGTGTATGCAAGACAAAGTTTCTTAAG 1337  
 DB 1561 GAGAGCCCAAAAAGATACAGAAATCGAATTTGATCTGGTGTATGCAAGACAAAGTTTCTTAAG 1619  
 QY 1338 CGAGCTACAGGAAACCTCAGCAGCTACAGAAAGAAATACCCCTAAAATGCTGAAAGAAATTTG 1397  
 DB 1620 CGAGCTACAGGAAACCTCAGCAGCTACAGAAAGAAATACCCCTAAAATGCTGAAATTTG 1678  
 QY 1398 GATTTCTCTTTGAAAGAGCTAAATTTTGTGAAAGATTTGGATTTATGCTCCAAACTTCAG 1457  
 DB 1679 GATTTCTCTTTGAAAGAGCTAAATTTTGTGAAAGATTTGGATTTATGCTCCAAACTTCAG 1738  
 QY 1458 ATGATCGAATGTCCAGGAAAGGTTTGAAGAGCCAAATGCTGTGTAATAGTGTAAATGAA 1517  
 DB 1739 ATGATCGAATGTCCAGGAAAGGTTTGAAGAGCCAAATGCTGTGTAATAGTGTAAATGAA 1798  
 QY 1518 CAATTTAAATACACACCGAGTATTCGTGATTTCAAGAAAAAACAATTTGAATGTGTTCCC 1577  
 DB 1799 CAATTTAAATACACACCGAGTATTCGTGATTTCAAGAAAAAACAATTTGAATGTGTTCCC 1858  
 QY 1578 GAAAAAGAACTTTTGTGCTGCTGTTTTTGTGTAAGTCAACGAAACAGCGGGAATTCATGCTTA 1637

```
Db 1859 GAAAGAGACCTTTCCTGCTGCTGTTTGTAGTCAACGAAACAGCGGGAAATCCATGCTTA 1918
Qy 1638 GAAGAGAACGACGTTGT-----TAA 1657
Db 1919 GAAGAGAACGACGTTGTAGTGTCTTCTACGTAGATTATTCGGAATAATTTTCAGTAA 1978
Qy 1658 GTTCTACACCGAATTAATTTGGTGTTCGAAAGTTCGGTGGATACGAATTTGGTGCATGA 1717
Db 1979 GTTCTACACCGAATTAATTTGGTGTTCGAAAGTTCGGTGGATACGAATTTGGTGCATGA 2038
Qy 1718 AAACAGAGGAGCGCAATCTATTATGTAGCGACGCAAGAAATGAATATGCC----- 1769
Db 2039 AAACAGAGGAGCGCAATCTATTATGTAGCGACGCAAGAAATGAATATGCCGTAAGTTT 2098
Qy 1770 -----TTCTACAAAAATTTGTACACTAAAT 1793
Db 2099 CAGAAAAATTGAAGTTTTTAAATATCATATTTACAGTTCTACAAAAATTTGTACACTAAAT 2158
Qy 1794 ACCGGAATCGGTAGATTTGAATAGCCGCAACAGAGCGGAAGATATGTTTGAACGTCCT 1853
Db 2159 ACCGGAATCGGTAGATTTGAATAGCCGCAACAGAGCGGAAGATATGTTTGAACGTCCT 2218
Qy 1854 CCGGATAAAGAAACAAAAAGTCTTAATGTTCAATTATCATTTTCCAAACGACAACTGAATGCT 1913
Db 2219 CCGGATAAAGAAAC-AAAAGTCTTAATGTTCAATTATCATTTTCCAAACGACAACTGAATGCT 2277
Qy 1914 TAGCGTTTGTGAAACATTAATTCGATCACACCATCGGTAGCTAATCAGCATATTTACT 1973
Db 2278 TAGCGTTTGTGAAACATTAATTCGATCACACCATCGGTAGCTAATCAGCATATTTACT 2337
Qy 1974 TCTGAAACAGTCAAAAAGCTTTGGCATCTAAGGACAGAGAAAGATCAAAAAGCAAT 2033
Db 2338 TCTGAAACAGTCAAAAAGCTTTGGCATCTAAGGACAGAGAAAGATCAAAAAGCAAT 2397
Qy 2034 TTCTATCAAAATTCGATTTGAAATCAACGCGAAATTTAGAGGTATTAACACGAGGCTTGAC 2093
Db 2398 TTCTATCAAAATTCGATTTGAAATCAACGCGAAATTTAGAGGTATTAACACGAGGCTTGAC 2457
Qy 2094 TGCTCAGAAATTCGAGAAATATCACCAAGAAAGAAAAAG-AAAAGCGGAAACAAATGCCATT 2152
Db 2458 TGCTCAGAAATTCGAGAAATATCACCAAGAAAGAAAAAGCAAGCAATGCCATT 2517
Qy 2153 AACTATGATATTTGGAAATGATGATTAATCACTCAACCTCTACAGTGGAAATGATTTATTC 2212
Db 2518 AACTATGATATTTGGAAATGATGATTAATCACTCAAC-CTCCTACAGTGGAAATGATTTATTC 2576
Qy 2213 TATAGCGGCTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGT 2272
Db 2577 TATAGCGGCTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGT 2636
Qy 2273 GACTCAAGAAAGATGTCGTCGCGGTGAGCGTGAGTGGCTCATGGACGGGAAAGAAACAGA 2332
Db 2637 GACTCAAGAAAGATGTCGTCGCGGTGAGCGTGAGTGGCTCATGGACGGGAAAGAAACAGA 2696
Qy 2333 TATTTTGGAAAGCAAGTTCGTGAAATTTGCTCAGAGAAATTCGCAGAA----- 2378
Db 2697 TATTTTGGAAAGCAAA-TTCGTGAAATTTGCTCAGAGAAATTCGCAGAAATGTCGTTGCTTGA 2755
Qy 2379 -----AACAACGACAATC 2391
Db 2756 GTATTTAAAGATCTCTGGATTTTAAATTTTGTAACTTTTGAACACGACAACTC 2815
Qy 2392 GAGCACCAAGCAATTTAGTGTCTATCGAGACGGAGTTAGCGATTCGGAGATGCTACGCTG 2451
Db 2816 GAGCACCAAGCAATTTAGTGTCTATCGAGACGGAGTTAGCGATTCGGAGATGCTACGCTG 2875
Qy 2452 TTAGTCATGATGAGCTTCGATCTTTTAAAGCGAAGTAAACAAATTCATGTCGGAAACGGG 2511
Db 2876 TTAGTCATGATGAGCTTCGATCTTTTAAAGCGAAGTAAACAAATTCATGTCGGAAACGGG 2935
Qy 2512 ATGGAAGATCCAGAGCGAAGTACACGTTCAATTCGTGATTCAGAAAGCAATACAC 2571
```

```
Db 2936 ATGGAAGATTCAGACCGAAGTACACGTTTCATTGTGATTTCAGAAAGACACAAATCAC 2995
Qy 2572 GATTGCTTCGAAGAAATGGAAGAAAGATGAAGCCAGTGGTCAATAAAGATCTTACTCTGCTG 2631
Db 2996 GATTGCTTCGAAGAAATGGAAGAAAGATGAAGCCAGT-GTCAATAAAGATCTTACTCTGCTG 3054
Qy 2632 AAACAGATGCTGCTGCTGCTGCTGTTAAACAAATGGGAGGAGGATATGAAGAAAGCAAG 2691
Db 3055 AAACAGATGCTGCTGCTGCTGCTGTTAAACAAATGGGAGGAGGATATGAAGAAAGCAAG 3114
Qy 2692 AAACGGAATTTGGAACCCCATCATCCGGAACAACTGTGGATAAATCTTCTGTTTCGAAT 2751
Db 3115 AAACGGAATTTGGAACCCCATCATCCGGAACAACTGTGGATAAATCTTCTGTTTCGAAT 3174
Qy 2752 ACAAAATTCGAATTTTCTTGGCATCTCATCATGCTGCTTCTTGGTACATCTCGTCCAGGAC 2811
Db 3175 ACAAAATTCGAATTTTCTTGGCATCTCATCATGCTGCTTCTTGGTACATCTCGTCCAGGAC 3234
Qy 2812 ATTACACTGTTATGTATGACGATAAAGGAATGAGCAGCAAGATGAAGTCTAT----- 2861
Db 3235 ATTACACTGTTATGTATGACGATAAAGGAATGAGCAGCAAGATGAAGTCTATGTTAAGCGTTT 3294
Qy 2862 ----- 2861
Db 3295 TGAATAGCAGTTAGCGATTTTAGGATTTTGTAAATCCGCATATAGTTATTATAAAAAAATG 3354
Qy 2862 -----AAAATGACCTACCGACTTTGCTTTTCTCTCTCTAGATGTCGAAAAACCCATCTCG 2915
Db 3355 TTTTCAGAAATGACCTACCGACTTTGCTTTTCTCTCTCTGCTAGATGTCGAAAAACCCATCTCG 3414
Qy 2916 TTGCTGTTCCGTTTCATATGCTCATATTCATGCTGAAAAAGCGAAAGAGCTTTTATCGA 2975
Db 3415 TTGCTGTTCCGTTTCATATGCTCATATTCATGCTGAAAAAGCGAAAGAGCTTTTATCGA 3474
Qy 2976 ACTTACAAGGAACATTTACATCCGCTGACTATGCACAGCCACGACTCGACAGAAATGGAA 3035
Db 3475 ACTTACAAGGAACATTTACATCCGCTGACTATGCACAGCCACGACTCGACAGAAATGGAA 3534
Qy 3036 CATTTTCTCAAACTAACCGTGAAGTACCCCTGGAATGTCGTCATAACTTTTGCAAAA 3095
Db 3535 CATTTTCTCAAACTAACCGTGAAGTACCCCTGGAATGTCGTCATAACTTTTGCAAAA 3594
Qy 3096 GTGTCGCCGTTTCAATCAAAATTTTTCATTTGATGATATTTGTTACTTTTAAAG 3155
Db 3595 GTGTCGCCGTTTCAATCAAAATTTTTCATTTGATGATATTTGTTACTTTTAAAG 3654
Qy 3156 CCGGTTTCAAAAATTCATTCCATGACTAAAGTTTTCATAAATTTACTTGAATTT 3210
Db 3655 CCGGTTTCAAAAATTCATTCCATGACTAAAGTTTTCATAAATTTACTTGAATTT 3709
```

## RESULT 3

ADBI7459

ID ADBI7459 standard; cDNA; 2827 BP.

XX ADBI7459;

AC ADBI7459;

XX 20-NOV-2003 (first entry)

XX Soybean post-transcriptional gene silencing 2wille homolog cDNA #1.

XX Soybean; ss; gene; post-transcriptional gene silencing; PTGS;

XX RNA degradation; transgene expression; plant; transgenic plant;

XX transposon activity; meristem activity; architecture; development;

XX proliferation; callus propagation.

XX Glycine max.

XX Key

XX Location/Qualifiers

XX 2..2467

XX /\*tag= a

XX /partial

XX /product= "2wille homolog"



FT /note= "No start codon shown"

XX US2003077623-A1.

XX 24-APR-2003.

XX 17-JUN-2002; 2002US-00174363.

XX 18-JUN-2001; 2001US-0298973P.

XX (BUTL/) BUTLER K H.

XX (CAHO/) CAHOON R E.

XX (HARV/) HARVELL L T.

XX (RAFA/) RAFALSKI J A.

XX (SAKA/) SAKAI H.

XX Butler KH, Cahoon RE, Harvell LT, Rafalski JA, Sakai H;

XX WPI: 2003-644827/61.

XX P-PSDB; ADS17460.

XX New polynucleotides encoding polypeptides comprising post-transcriptional

XX gene silencing activity useful in creating transgenic plants having

XX enhanced or deficient post-transcriptional gene silencing.

XX Disclosure; Page 21-22; 107pp; English.

XX The invention discloses an isolated polynucleotide encoding a polypeptide

XX having post-transcriptional gene silencing (PTGS) activity. PTGS operates

XX at the level of sequence specific RNA degradation and down regulates

XX transgene expression in plants. Also claimed is a recombinant DNA

XX construct comprising the polynucleotide operably linked to at least one

XX regulatory sequence, transforming a cell with the polynucleotide, a cell

XX comprising the recombinant construct, producing a polypeptide having post

XX -transcriptional gene silencing activity, producing a plant, a plant and

XX a seed comprising the recombinant DNA construct and an isolated

XX polypeptide comprising post-transcriptional gene silencing activity. Also

XX disclosed are methods for obtaining a nucleic acid fragment encoding all,

XX or a substantial portion of, the amino acid sequence encoding a

XX polypeptide involved in post-transcriptional gene silencing, positive

XX selection of a transformed cell and altering the level of expression of a

XX polypeptide involved in post-transcriptional gene silencing in a host

XX cell. The polynucleotides encoding polypeptides involved in gene

XX silencing are useful in altering the post-transcriptional gene silencing

XX pathway in plant cells. These are specifically useful in creating

XX transgenic plants where the levels of polypeptides involved in post-

XX transcriptional gene silencing are altered with respect to non-transgenic

XX plants which would result in plants with an enhanced or deficient post-

XX transcriptional gene silencing. The polynucleotides may also be used to

XX manipulate transposon activity, meristem activity, plant architecture and

XX development, proliferation of undifferentiated plant cells in culture for

XX callus propagation, as probes for genetically and physically mapping the

XX genes that they are part of and as markers for traits linked to those

XX genes. The sequence presented is a cDNA encoding a polypeptide having

XX post-transcriptional gene silencing activity.

XX Sequence 2827 BP; 876 A; 509 C; 645 G; 797 T; 0 U; 0 Other;

SQ

Query Match 2.0%; Score 63.8; DB 9; Length 2827;

Best Local Similarity 52.0%; Pred. No. 0.0002;

Matches 143; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 2674 ATATGAAGAAGCAAGAACTGGAAATGTGAACCCATCATCCGGAACACTGTGGATA 2733

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2028 ACAGGGACAGAGACAGACAGATCGGAGTGGGAATATATTCCTGGGACTGTGTGTGATA 2087

QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2734 AACTTATGCTTTCGAATACAAATCTGATTTTCTTGGCATCTCATGTGCTCTTG 2793

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2088 CCAAAATCTGCCATCCACAGAATTTGATTTTATCTCTGCAGCCATGTGGATCAGG 2147

QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2794 GTACATCTCGTCCAGACATTACACTGTTATGTATGACGATAAAGGAATGAGCAAGATG 2853

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2148 GTACTAGTCGGCCAGCTCATTTATCATGTCCTGTGGGATGAAACAACTTCACACCTGATG 2207

QY 2854 AGTCTATAAATGACCTAGCGACTTGCTTTTCTCTGCTAGATGTCGAAACCATCT 2913

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2208 GAAATTCAGTCTCTGACAAACAAACCTTTGTTATACATATGCCAGGTGTACACGCTCAGTAT 2267

QY 2914 CGTTGCCTGTTCCGGTTCATTATGCTCAATTATCA 2948

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

2268 CAGTTGTTCTCTCCAGCATATATTATGACATTTAGCA 2302

RESULT 4

AEC75683

ID AEC75683 standard; cDNA; 2827 BP.

XX AEC75683;

XX 17-NOV-2005 (first entry)

XX Soybean Zwillle (ZLL) homolog FIS cDNA clone, SEQ ID NO: 3.

XX Gene silencing; transgenic plant; cell proliferation; agriculture;

XX crop improvement; Zwillle protein; gene; ss.

XX Glycine max.

XX Key Location/Qualifiers

XX CDS 2..2467

XX FT /\*tag= a

XX FT /product= "soybean Zwillle homolog protein"

XX FT /partial

XX FT /note= "No start codon"

XX US2005204427-A1.

XX 15-SEP-2005.

XX 30-MAR-2005; 2005US-00093888.

XX 18-JUN-2001; 2001US-0298973P.

XX 17-JUN-2002; 2002US-00174363.

XX (BUTL/) BUTLER K H.

XX (CAHO/) CAHOON R E.

XX (HARV/) HARVELL L T.

XX (RAFA/) RAFALSKI J A.

XX (SAKA/) SAKAI H.

XX Butler KH, Cahoon RE, Harvell LT, Rafalski JA, Sakai H;

XX WPI: 2005-618219/63.

XX P-PSDB; AEC75684.

XX New polynucleotide encoding a polypeptide having post-transcriptional

XX gene silencing activity, useful for manipulating plant architecture and

XX development, or for propagating callus.

XX Disclosure; SEQ ID NO 3; 107pp; English.

XX The invention relates to a polynucleotide encoding a polypeptide having

XX post-transcriptional gene silencing (PTGS) activity. The invention also

XX relates to a method for producing transgenic plants. The sequences and

XX method of the invention are useful for manipulating transposon activity,

XX meristem activity, plant architecture and development or proliferation of

XX undifferentiated plant cells in culture and for propagating callus. The

XX present sequence is the soybean Zwillle (ZLL) homolog cDNA. This sequence

XX is involved in PTGS activity.

XX Sequence 2827 BP; 876 A; 509 C; 645 G; 797 T; 0 U; 0 Other;

SQ

Query Match 2.0%; Score 63.8; DB 14; Length 2827;

Best Local Similarity 52.0%; Pred. No. 0.0002;

Matches 143; Conservative 0; Mismatches 132; Indels 0; Gaps 0;



PT Novel human genes and their expression products which are differentially  
 PT expressed in different cell types.

PS Claim 1; Page 1234; 2479pp; English.

XX The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AA212532 to AA21779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one of  
 CC the 5248 polynucleotide sequences given in AA212532 to AA21779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purposes, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists

XX Sequence 734 BP; 219 A; 159 C; 155 G; 193 T; 0 U; 8 Other;

Query Match 1.9%; Score 62.4; DB 2; Length 734;

Best Local Similarity 51.8%; Pred. No. 0.00027;

Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 2698 GAATTTGTAACCCATCATCCGGAACAACTGTGGATAAACTTATCGTTTCGAAATACAAAT 2757

DB 209 GAAGTGGCAATATCCAGCTGGAAACAAAGTTGATACAGACATTACACACCCATATGACT 268

QY 2758 TCGATTTTTCTTGGCATCTCATCTGCTGGTACATCTCGTCCAGACATTACA 2817

DB 269 TCGATTTTTTACCTCTGTAGCCATGCTGGAAATACAGGGTACCGAGTCTCTTACACACTATC 328

QY 2818 CTGTTATGTATGACGATAAAGGAATGAGCCAGATGAAGTCTATAAAATGACCTACGGAC 2877

DB 329 ATGTTTATGGATGATTAAGTCTTTACTGAGATGAAGTCTTACGCTGCTAACTTACCAGC 388

QY 2878 TTGCTTTTCTCTCTGCTAGATGTCGAAACCCATCTGTTGCTGCTGTTCCGGTTCATTATG 2937

DB 389 TCTGCCACACTTACGTCACGCTGTACACGATCTGTTTCTATACCTGCACACAGCGTATTATG 448

QY 2938 CTCATTATCATGTGAAAAGCGAAGAGCTT 2969

DB 449 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 480

RESULT 7

ADJ94706

ID ADJ94706 standard; cDNA; 2568 BP.

XX AC ADJ94706;

XX 06-MAY-2004 (first entry)

XX Human eukaryotic translation initiation factor 2C4 (eIF2C4) cDNA.

XX cytosolic; antiinflammatory; virucide; immunosuppressive; tumour;

XX inflammatory; infectious disease; viral infection; degenerative;

XX autoimmune; gene therapy; Argonaute family;

XX eukaryotic translation initiation factor 2C4; eIF2C4; human; ss; gene.

XX Homo sapiens.

XX WO2004007718-A2.

XX

PD 22-JAN-2004.

XX 10-JUL-2003; 2003WO-EP007516.

XX 10-JUL-2002; 2002EP-00015532.

PR 23-AUG-2002; 2002EP-00018906.

XX (PLAC ) MAX PLANCK GRS FOERDERUNG WISSENSCHAFTEN.

PA Tuschl T, Martinez J, Patkaniowska A, Urlaub H, Luehrmann R;

PI WPI; 2004-122948/12.

XX P-PSDB; ADJ94700.

XX New single-stranded RNA molecule having a length from 14-50 nucleotides,  
 PT useful for preventing or treating tumor, inflammatory, infectious, e.g.  
 PT viral infections, degenerative and autoimmune diseases.

XX Example; Fig 16; 73pp; English.

XX The invention relates to a novel single-stranded RNA molecule having a  
 CC length from 14-50 nucleotides where at least 14-20 of the 5'-most  
 CC nucleotides are substantially complementary to a target transcript. The  
 CC RNA molecule of the invention demonstrates cytostatic, antiinflammatory,  
 CC virucide and immunosuppressive activities and may be useful for  
 CC inhibiting the expression of a target gene in vitro or in vivo,  
 CC preferably for preventing or treating diseases associated with the  
 CC overexpression of at least one target transcript. The diseases may be  
 CC selected from tumour diseases, inflammatory diseases and autoimmune diseases.  
 CC such as viral infections, degenerative diseases and autoimmune diseases.  
 CC Furthermore, the molecules of the invention may be utilised during gene  
 CC therapy. The current sequence is that of the human eukaryotic translation  
 CC initiation factor 2C4 (eIF2C4) cDNA of the invention.

XX Sequence 2568 BP; 753 A; 568 C; 600 G; 647 T; 0 U; 0 Other;

Query Match 1.9%; Score 62.4; DB 12; Length 2568;

Best Local Similarity 51.8%; Pred. No. 0.00042;

Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 2698 GAATTTGTAACCCATCATCCGGAACAACTGTGGATAAACTTATCGTTTCGAAATACAAAT 2757

DB 2165 GAAGTGGCAATATCCAGCTGGAAACAAAGTTGATACAGACATTACACACCCATATGACT 2224

QY 2758 TCGATTTTTCTTGGCATCTCATCTGCTGGTACATCTCGTCCAGGACATTACA 2817

DB 2225 TCGATTTTTTACCTCTGTAGCCATGCTGGAAATACAGGGTACCGAGTCTCTTACACTATC 2284

QY 2818 CTGTTATGTATGACGATAAAGGAATGAGCCAGATGAAGTCTATAAAATGACCTACGGAC 2877

DB 2285 ATGTTTATGGATGATTAAGTCTTTACTGAGATGAAGTCTTACGCTGCTAACTTACCAGC 2344

QY 2878 TTGCTTTTCTCTCTGCTAGATGTCGAAACCCATCTGTTGCTGCTGTTCCGGTTCATTATG 2937

DB 2345 TCTGCCACACTTACGTCACGCTGTACACGATCTGTTTCTATACCTGCACACAGCGTATTATG 2404

QY 2938 CTCATTATCATGTGAAAAGCGAAGAGCTT 2969

DB 2405 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 2436

RESULT 8

AAK53425

ID AAK53425 standard; cDNA; 2914 BP.

XX AC AAK53425;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 3938.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;



CC anyotropic lateral sclerosis, stroke, immune deficiencies (e.g. human  
 CC immunodeficiency virus (HIV), severe combined immunodeficiency or  
 CC infections), autoimmune disorders (e.g. rheumatoid arthritis, Guillain-  
 CC Barre syndrome or graft-versus-host disease), cancers (e.g. thyroid  
 CC cancer, lung cancers, small cell carcinoma, Kaposi's sarcoma, brain  
 CC tumours, prostate cancer, ovarian cancer or leukaemia), or inflammatory  
 CC disorders (e.g. inflammatory bowel disease or Crohn's disease). The  
 CC polynucleotides and proteins are useful for screening peptides or small  
 CC molecule inhibitors and agonists that are useful for treating these  
 CC diseases. The polypeptide is also useful as molecular markers, or as a  
 CC food supplement. The present sequence is one of the human polynucleotides  
 CC of the invention encoding a cytokine or cell proliferation/  
 CC differentiation-related protein

XX SQ Sequence 2914 BP; 856 A; 638 C; 668 G; 752 T; 0 U; 0 Other;

Query Match 1.9%; Score 62.4; DB 8; Length 2914;  
 Best Local Similarity 51.8%; Pred. No. 0.00044;  
 Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 2698 GAATTGTGAACCCATCATCCGGAACAACTGTGGATAAATTATCGTTTCGAAATACAAAT 2757  
 DB 2180 GAAGTGGCAATATCCAGCTGGAAACAAAGTTGATACAGACATTACACACCCATATGAGT 2239  
 QY 2758 TCGATTTTCTTGGCATCTCATCGTGTCTTGGTACATCTCGTCCAGGACATTACA 2817  
 DB 2240 TCGATTTTACCTCTGTAGCCATCGCTGGAATACAGGGTACCACTCGTCTTCCACACTATC 2299  
 QY 2818 CTGTTATGTATGACGATAAAGGAATGAGCCAGCAAGTCTATAAAATGACCTACGGAC 2877  
 DB 2300 ATGTTTATGGATGATTAAGTCTTACTGTCAGATGAATCTTACGCTGAATACCAAGC 2359  
 QY 2878 TTGCTTTTCTCTGTAGATGTCGAAACCCATCTCGTTGCTGTTCGGTTTCATTATG 2937  
 DB 2360 TCTGCCACACTTACGTACGCTGTACACGATCTGTTTCTATACCTGACCAAGCGTATTATG 2419  
 QY 2938 CTCATTATCATGTGAAAGCGAAGAGCTT 2969  
 DB 2420 CTCACCTGTAGCATTTAGAGCCAGATATCAT 2451

RESULT 10  
 AAH16058  
 ID AAH16058 standard; cDNA; 3050 BP.

XX AC AAH16058;  
 XX 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:14738.  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX Homo sapiens.  
 XX EP1074617-A2.  
 XX 07-FEB-2001.  
 XX 28-JUL-2000; 2000BP-00116126.  
 XX 29-JUL-1999; 99JP-00248036.  
 XX 27-AUG-1999; 99JP-00300253.  
 XX 11-JAN-2000; 2000JP-00118776.  
 XX 02-MAY-2000; 2000JP-00183767.  
 XX 09-JUN-2000; 2000JP-00241899.  
 XX (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;  
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.

XX Claim 8; SEQ ID NO 14738; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention

XX SQ Sequence 3050 BP; 886 A; 687 C; 702 G; 775 T; 0 U; 0 Other;

Query Match 1.9%; Score 62.4; DB 4; Length 3050;  
 Best Local Similarity 51.8%; Pred. No. 0.00045;  
 Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 2698 GAATTGTGAACCCATCATCCGGAACAACTGTGGATAAATTATCGTTTCGAAATACAAAT 2757  
 DB 2280 GAAGTGGCAATATCCAGCTGGAAACAAAGTTGATACAGACATTACACACCCATATGAGT 2339  
 QY 2758 TCGATTTTCTTGGCATCTCATCGTGTCTTGGTACATCTCGTCCAGGACATTACA 2817  
 DB 2340 TCGATTTTACCTCTGTAGCCATCTGGAATACAGGGTACCACTCGTCTTCCACTATC 2399  
 QY 2818 CTGTTATGTATGACGATAAAGGAATGAGCCAGATGAAGTCTATAAAATGACCTACGGAC 2877  
 DB 2400 ATGTTTATGGATGATTAAGTCTTACTGTCAGATGAATCTTACCTGCTGAATACCAAGC 2459  
 QY 2878 TTGCTTTTCTCTGTAGATGTCGAAACCCATCTCGTTGCTGTTCGGTTTCATTATG 2937  
 DB 2460 TCTGCCACACTTACGTACGCTGTACAGATCTGTTTCTATACCTGACCAAGCGTATTATG 2519  
 QY 2938 CTCATTATCATGTGAAAGCGAAGAGCTT 2969  
 DB 2520 CTCACCTGTAGCATTTAGAGCCAGATATCAT 2551

RESULT 11  
 ADH77114  
 ID ADH77114 standard; cDNA; 3050 BP.

XX ADH77114;  
 XX 22-APR-2004 (first entry)  
 XX Human PAZ/PIWI domain-containing protein cDNA.  
 XX cytosolic; PAZ/PIWI domain-containing protein inhibitor;  
 KW PAZ/PIWI domain-containing protein; hyperproliferative disorder; cancer;  
 KW aberrant cellular differentiation; human;  
 XX PAZ/PIWI domain-containing protein; gene; ss.  
 XX

OS Homo sapiens.  
 PN US2003232442-A1.  
 XX 18-DEC-2003.  
 PD 17-JUN-2002; 2002US-00175492.  
 PF 17-JUN-2002; 2002US-00175492.  
 PR (ISIS-) ISIS PHARM INC.  
 XX Dobie KW;  
 XX WPI; 2004-052174/05.  
 DR P-PSDB; ADH77275.  
 XX  
 PT New antisense oligonucleotide targeted to a nucleic acid encoding a  
 PT PAZ/PIWI domain-containing protein, useful for treating cancer or a  
 PT disease arising from aberrant cellular differentiation.  
 XX  
 PS Example 13; SEQ ID NO 4; 119pp; English.  
 XX  
 CC The invention describes a compound 8-80 nucleobases in length targeted  
 CC to, and which specifically hybridizes with a nucleic acid molecule  
 CC encoding a PAZ/PIWI domain-containing protein, and inhibits the  
 CC expression of a PAZ/PIWI domain-containing protein. The compound,  
 CC associated with PAZ/PIWI domain-containing protein, such as a  
 CC hyperproliferative disorder e.g. cancer, or a disease or condition  
 CC arising from aberrant cellular differentiation. They are also useful in  
 CC research and diagnostics for modulating the expression of PAZ/PIWI domain  
 CC -containing protein. This sequence encodes a human PAZ/PIWI domain-  
 CC containing protein.  
 XX  
 SQ Sequence 3050 BP; 886 A; 687 C; 702 G; 775 T; 0 U; 0 Other;  
 Query Match 1.9%; Score 62.4; DB 12; Length 3050;  
 Best Local Similarity 51.8%; Pred. No. 0.00045;  
 Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 2698 GAATTGTGAACCCATCATCCGGAACAACCTGGGATAAACTTATCGTTTCGAAATACAAAT 2757  
 DB 2280 GAAGTGCATATCCGAGCTGGAAACAACAGTTGATACAGACATTACACACCATATGAT 2339  
 QY 2758 TCGATTTTCTTGGCATCTCATCGTGTCTTGGTACATCTCGTCCAGGACATTACA 2817  
 DB 2340 TCGATTTTACCTCTGTAGCCATCTCGAATACAGGTACAGTCTGCTTCACACTATC 2399  
 QY 2818 CTGTTATGTATGACGATAAGGATGAGCCAGATGAAGTCTATAAATGACCTACGGAC 2877  
 DB 2400 ATGTTTATGGATGATACTGCTTTTACTCGAGATGAACCTTCAGCTGCTAACTTACCAGC 2459  
 QY 2878 TTGCTTTTCTCTGCTAGATGTCGAAACCCATCTCGTTCCTGCTTCCGTTTCATTATG 2937  
 DB 2460 TCTGCCACACTTACGTAGCTGTACAGATCTGTTTCTATACCTTGCACCGGTTATATG 2519  
 QY 2938 CTCATTATCATGTGAAAGGCGAAGAGCTT 2969  
 DB 2520 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 2551  
 RESULT 12  
 AAD55860  
 ID AAD55860 standard; cDNA; 3325 BP.  
 XX  
 AC AAD55860;  
 XX  
 DT 07-AUG-2003 (first entry)  
 XX Human nucleic acid associated protein (NAAP)-31 cDNA.  
 DE Human; nucleic acid associated protein; NAAP; stroke; AIDS; nootropic;  
 XX

KW cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;  
 KW developmental disorder; antiinflammatory; neuroprotective; thymimetic;  
 KW Cushing's syndrome; infection; gene therapy; cytostatic; anticonvulsant;  
 KW cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;  
 KW gene; ss.  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 CDS 497..2947  
 FT /\*tag= a  
 FT /product= "Human NAAP protein"  
 XX  
 PN WO2003006618-A2.  
 XX 23-JAN-2003.  
 PD 10-JUL-2002; 2002WO-US021971.  
 PF 12-JUL-2001; 2001US-0305089P.  
 XX 12-JUL-2001; 2001US-0305104P.  
 PR 13-JUL-2001; 2001US-0305325P.  
 PR 13-JUL-2001; 2001US-0305390P.  
 PR 19-JUL-2001; 2001US-0306960P.  
 PR 20-JUL-2001; 2001US-030694P.  
 PR 27-JUL-2001; 2001US-0308170P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;  
 PI Yue H, Baughn MR, Emerling BM, Lal PG, Lu DAM, Forsythe LJ;  
 PI Ramkumar J, Li JX, Becha SD, Duggan BM, Sanjanwala MM, Lee EA;  
 PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yao MG;  
 PI Barroso I, Tran B, Wallia NK, Hafalia AJA, Nguyen DB, Lu Y;  
 PI Arvizu CS;  
 XX  
 DR WPI; 2003-221732/21.  
 XX P-PSDB; AAE37044.  
 XX  
 PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
 PT treating and preventing diseases or conditions associated with the  
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
 PT infections.  
 XX  
 PS Claim 5; Page 255-256; 260pp; English.  
 XX  
 CC The invention relates to human nucleic acid associated proteins (NAAP)  
 CC and their corresponding nucleic acid sequences. The invention is useful  
 CC in diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression or overexpression of NAAP, such as cell  
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
 CC disorders, or infections. It is also useful in assessing the effects of  
 CC exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of NAAP. The NAAP or its fragments are useful in screening  
 CC compounds for which acts as their agonist or antagonist. The microarray  
 CC is useful in monitoring or measuring protein-protein interactions, drug-  
 CC target interactions, and gene expression profiles. NAAP DNA is used in  
 CC gene therapy. The present sequence is human NAAP cDNA  
 XX  
 SQ Sequence 3325 BP; 902 A; 809 C; 813 G; 801 T; 0 U; 0 Other;  
 Query Match 1.9%; Score 62.4; DB 8; Length 3325;  
 Best Local Similarity 51.8%; Pred. No. 0.00046;  
 Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 2698 GAATTGTGAACCCATCATCCGGAACAACCTGGGATAAACTTATCGTTTCGAAATACAAAT 2757  
 DB 2544 GAAGTGCATATCCGAGCTGGAAACAACAGTTGATACAGACATTACACACCATATGAT 2603  
 QY 2758 TCGATTTTCTTGGCATCTCATCGTGTCTTGGTACATCTCGTCCAGGACATTACA 2817

Db 2604 TCGATTTTACCTCTGTCAGCATGCTGGATAACAGGGTACCAGTCTGCTTACACTATC 2663  
Qy 2818 CRTGTTATGATGACGATAAGGAATGAGCGCAAGATGAAGTCTATAAAATGACCTACGGAC 2877  
Db 2664 ATGTTTTATGGGATGATACTGCTTTTACTGCAGATGAACCTTCAGCTGCTAACTTACCAGC 2723  
Qy 2878 TTGCTTTTCTCTGCTAGATGTCGAAACCCATCTGTTGCTGTTCGGTTTCATTATG 2937  
Db 2724 TCTGCCACACTTACGTACGCTGTCACGATCTGTTTCTATACCTGCACCGGATATTATG 2783  
Qy 2938 CTCATTATCATGTGAAAAGCGAAGAGCTT 2969  
Db 2784 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 2815

RESULT 13  
ABX34752  
ID ABX34752 standard; cDNA; 3736 BP.  
XX AC ABX34752;  
XX  
DT 13-FEB-2003 (first entry)  
XX  
DE Human mddt cDNA SEQ ID 313.  
XX  
KW MDDT; human; disease detection and treatment molecule polypeptide;  
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
KW haemostatic; nephrotropic; antianaemic; antipsoriasis; hepatotropic;  
KW gene therapy; protein replacement therapy; cell proliferative disorder;  
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;  
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
KW psoriasis; hepatitis; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200279449-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-US009944.  
XX  
PR 28-MAR-2001; 2001US-0279619P.  
PR 29-MAR-2001; 2001US-0280067P.  
PR 29-MAR-2001; 2001US-0280068P.  
PR 16-MAY-2001; 2001US-0291280P.  
PR 17-MAY-2001; 2001US-0291829P.  
PR 17-MAY-2001; 2001US-0291849P.  
PR 19-JUN-2001; 2001US-0299428P.  
PR 20-JUN-2001; 2001US-0299776P.  
PR 20-JUN-2001; 2001US-0300001P.  
XX  
(INCY-) INCYTE GENOMICS INC.  
XX  
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
PI Dufour GE, Hillman JR, Yu JY, Tuason O, Yap PE, Anshey SR;  
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;  
XX  
DR WPI; 2003-058431/05.  
DR P-PSDB; ABU11762.  
XX  
XX New purified disease detection and treatment molecule proteins and  
PT polynucleotides, useful for diagnosing, treating or preventing cancers  
PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
PT or hepatitis.  
XX  
PS Claim 1; SEQ ID NO 313; 339pp + Sequence Listing; English.  
XX  
XX This invention describes a novel disease detection and treatment molecule  
CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,  
CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,  
CC

CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides  
CC and the polypeptides of the invention can be used for gene therapy, of  
CC protein replacement therapy and are useful for treating a variety of  
CC diseases or conditions. These polypeptides or polynucleotides are  
CC particularly useful for diagnosing, treating or preventing cell  
CC proliferative disorders (e.g. cancers including adenocarcinoma,  
CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or  
CC hepatitis. ABX34440-ABX34835 encode the MDDT polypeptides represented in  
CC ABU11450-ABU11845, described in the disclosure of the invention. NOTE:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3736 BP; 1061 A; 853 C; 809 G; 1013 T; 0 U; 0 Other;  
Query Match 1.9%; Score 62.4; DB 8; Length 3736;  
Best Local Similarity 51.8%; Pred. No. 0.00048;  
Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
Qy 2698 GAATTGTGAACCCATCATCCGGAACAACCTGTGGATAAACTTATCGTTTGAATAACAAT 2757  
Db 2445 GAAGTGGCAATATCCAGCTGGAAACAACAGTTTGATACAGACATTACACACCCATATGAGT 2504  
Qy 2758 TCGATTTTCTTGGCAATCTCATCTGCTTGGTACATCTCGTCCAGGACATTACA 2817  
Db 2505 TCGATTTTTCACCTCTGTAGCCATGTGGAAATACAGGGTACCAGTCTCTTACACTATC 2564  
Qy 2818 CTGTTATGTATCAGCATAAAGGAATGAGCAAGATGAAGTCTATAAAATGACCTACGGAC 2877  
Db 2565 ATGTTTATGGATGATATACTGCTTTACTGCAAGTGAACCTCAGCTGCTAACTTACCAGC 2624  
Qy 2878 TTGCTTTTCTCTGCTAGATGTCGAAAACCCATCTCGTTGCTGTTCGGTTCATATG 2937  
Db 2625 TCTGCCACACTTACGTACGCTGTACAGATCTGTTTCTATACCTGCACCGCGTATTATG 2684  
Qy 2938 CTCATTATCATGTGAAAAGCGAAGAGCTT 2969  
Db 2685 CTCACCTGGTAGCATTTTAGAGCCAGATATCAT 2716  
RESULT 14  
ADQ80688  
ID ADQ80688 standard; DNA; 2679 BP.  
XX AC ADQ80688;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Arabidopsis thaliana TFL1-binding protein coding sequence #2.  
XX  
KW TFL1-binding protein; plant growth control; biotechnology;  
KW fishing industry; screening; gene; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2679  
FT /\*tag= a  
FT /product= "Arabidopsis thaliana TFL1-binding protein #2"  
XX  
PN JP2004208572-A.  
XX  
PD 29-JUL-2004.  
XX  
PF 27-DEC-2002; 2002JP-00381220.  
XX  
PR 27-DEC-2002; 2002JP-00381220.  
XX  
PA (DOKU-) DOKURITSU GYOSEI HOJIN KAGAKU GIJUTSU SH.  
XX WPI; 2004-538191/52.  
DR





Db 1091 CATTGGCAGCTTACAGAGC 1110

Search completed: July 5, 2006, 21:52:46  
Job time : 1730 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	3227	100.0	3227	9	US-10-645-746-2
2	3227	100.0	3227	10	US-10-645-735-2
3	3227	100.0	3227	16	US-11-144-985-2
4	2601	80.6	3719	9	US-10-645-746-1
5	2601	80.6	3719	10	US-10-645-735-1
6	2601	80.6	3719	16	US-11-144-985-1
7	63.8	2.0	2827	6	US-10-174-363-3
8	63.8	2.0	2827	13	US-11-093-888-3
9	62.4	1.9	490	3	US-09-918-995-10623
10	62.4	1.9	734	10	US-10-779-543-3372
11	62.4	1.9	2914	3	US-09-774-434-2
12	62.4	1.9	3050	7	US-10-175-492-4
13	62.4	1.9	3325	10	US-10-483-505-66
14	62.4	1.9	3578	10	US-10-756-149-37
15	60	1.9	1501	6	US-10-174-363-5
16	60	1.9	1501	13	US-11-093-888-5
17	58.4	1.8	970	8	US-10-424-599-43001

61 GTCA TTCATCTGATCCGAGATG

Qy 61 GTCATTCTCTCGATCCGGAGATGAAATGGCTTGGAGGCCCACTGGTAAATGCGACGGCA 120





|||||  
661 ATTTTGTGACGATATATATCTATTTCTGGAGTCTCTGAAATCGTTTTCACGATCCAAACA 720  
QY  
721 GATTCGAACAATCATTTAGAGTAGCACCAAGAATCGAAGCATGGTTTGGAAATTTACATTG 780  
Db  
721 GATTCGAACAATCATTTAGAGTAGCACCAAGAATCGAAGCATGGTTTGGAAATTTACATTG 780  
QY  
781 GAATCAAGAANTTTGTCGATGGTGAACCTGTGCTCAATTTTGCANATGTCGATTAACATAT 840  
Db  
781 GAATCAAGAANTTTGTCGATGGTGAACCTGTGCTCAATTTTGCANATTTGTCGATTAACATAT 840  
QY  
841 TCTACAATGCACCGAAAAATGTTCTTCTCTGGATTTATCTTCTCTAAATTTGTGCAACCCCACT 900  
Db  
841 TCTACAATGCACCGAAAAATGTTCTTCTCTGGATTTATCTTCTCTAAATTTGTGCAACCCCACT 900  
QY  
901 CGTGTAAACGATGATGTACGAAAGATCTTTAAACAAAACTGATGCGGGGAAAAAATGACAA 960  
Db  
901 CGTGTAAACGATGATGTACGAAAGATCTTTAAACAAAACTGATGCGGGGAAAAAATGACAA 960  
QY  
961 TCAGACAGCCGCGCCCAAGNATTCGACAAATTTTGGAAAAATTTGAAGCTGGAATGCG 1020  
Db  
961 TCAGACAGCCGCGCGCCCAAGNATTCGACAAATTTTGGAAAAATTTGAAGCTGGAATGCG 1020  
QY  
1021 CAGAAGTTTGGGATACGAAATGTTGAGATTCGACAGAACGACATCTGACATTTCTAGATT 1080  
Db  
1021 CAGAAGTTTGGGATACGAAATGTTGAGATTCGACAGAACGACATCTGACATTTCTAGATT 1080  
QY  
1081 TGTGCGAGAAAACTCTCTTGTATATAAAGTCACTGTTAAATCGGACAGAGGAAGAAATG 1140  
Db  
1081 TGTGCGAGAAAACTCTCTTGTATATAAAGTCACTGTTAAATCGGACAGAGGAAGAAATG 1140  
QY  
1141 CAAAAAAGTACGATACATTTGTTCAAAATCTATGAGGAAAAACAAAAAGTTCAATTTGAGT 1200  
Db  
1141 CAAAAAAGTACGATACATTTGTTCAAAATCTATGAGGAAAAACAAAAAGTTCAATTTGAGT 1200  
QY  
1201 TTCCCCACCTTACCACTAGTCAAAAGTTAAAGTGGAGCAAAAGAAATACGCTGTACCAATGG 1260  
Db  
1201 TTCCCCACCTTACCACTAGTCAAAAGTTAAAGTGGAGCAAAAGAAATACGCTGTACCAATGG 1260  
QY  
1261 AACATCTTGAAGTTCTATGAGAGCCCAAAAGATACAAAGATTCGAATTTGATCTGGTGATGC 1320  
Db  
1261 AACATCTTGAAGTTCTATGAGAGCCCAAAAGATACAAAGATTCGAATTTGATCTGGTGATGC 1320  
QY  
1321 AAGACAAGTTTCTAAAGCGAGCTACACGAAACCTCACGACTACAAAGAAAAATACCCCTAA 1380  
Db  
1321 AAGACAAGTTTCTAAAGCGAGCTACACGAAACCTCACGACTACAAAGAAAAATACCCCTAA 1380  
QY  
1381 AAATGCTGAAAGAAATTTGGATTTCTTCTGAGAGCTAAATTTTGTGAAAGATTTGGAT 1440  
Db  
1381 AAATGCTGAAAGAAATTTGGATTTCTTCTGAGAGCTAAATTTTGTGAAAGATTTGGAT 1440  
QY  
1441 TATGCTCCAAAATTTCAGATGATGCAATGTCGAGGAAAGGTTTTGAAAGGCCAATGCTTTG 1500  
Db  
1441 TATGCTCCAAAATTTCAGATGATGCAATGTCGAGGAAAGGTTTTGAAAGGCCAATGCTTTG 1500  
QY  
1501 TGAATAGTCTAAATGAAACAAATTTAAATGACACCAAGTGAATTCGTGATTTCAAGAAAAAC 1560  
Db  
1501 TGAATAGTCTAAATGAAACAAATTTAAATGACACCAAGTGAATTCGTGATTTCAAGAAAAAC 1560  
QY  
1561 AAATGGAATGTTGTTCCGAAAAAGAACTTTGCTGCTGCTTTTGTAGTCAACGAAACAG 1620  
Db  
1561 AAATGGAATGTTGTTCCGAAAAAGAACTTTGCTGCTGCTTTTGTAGTCAACGAAACAG 1620  
QY  
1621 CGGGAAATCCATGCTTTAGAGAGAACGAGTTGTTAAGTTCTACACCGAACTAAATTTGGTG 1680  
Db  
1621 CGGGAAATCCATGCTTTAGAGAGAACGAGTTGTTAAGTTCTACACCGAACTAAATTTGGTG 1680  
QY  
1681 GTTGAAGTTCCGTGGAATACGAATTTGGTGCCAAATGAAACAGAGAGCGCAATCTATTTA 1740  
Db  
1681 GTTGAAGTTCCGTGGAATACGAATTTGGTGCCAAATGAAACAGAGAGCGCAATCTATTTA 1740  
QY  
1741 TGTACGACGCGACGAAAAAATGAATATGCTTCTCAAAAAATTTGTACATAAATACCGGAA 1800

1741 TGTACGACGCGACGAAAAAATGAATATGCTTCTCAAAAAATTTGTACACTAAATACCGGAA 1800  
QY  
1801 TCGGTAGATTTCGAAATAGCCGCAACAGAAAGCGAAGNATATGTTGNAACGCTCTCCCGATA 1860  
Db  
1801 TCGGTAGATTTCGAAATAGCCGCAACAGAAAGCGAAGNATATGTTGNAACGCTCTCCCGATA 1860  
QY  
1861 AAGAACAAAAAGTCTTAAATGTTTCAATTTATCATTTCCAAACGCAAACTGAATGCTTACGGTT 1920  
Db  
1861 AAGAACAAAAAGTCTTAAATGTTTCAATTTATCATTTCCAAACGCAAACTGAATGCTTACGGTT 1920  
QY  
1921 TTGTGAAACATTTATTCGATCACACATCGGTGTAGCTTAATCAGCATATTTACTTTGAAA 1980  
Db  
1921 TTGTGAAACATTTATTCGATCACACATCGGTGTAGCTTAATCAGCATATTTACTTTCTGAAA 1980  
QY  
1981 CAGTCACAAAAAGCTTTGGCATCACTAAGGACAGGAAAGGATCAAAACGAAATTTCTATC 2040  
Db  
1981 CAGTCACAAAAAGCTTTGGCATCACTAAGGACAGGAAAGGATCAAAACGAAATTTCTATC 2040  
QY  
2041 AAATTTGCAATTTGAAAAATCAACGCGAAAATTAGGAGGTATTAAACGAGAGCTTTGACTGCTCAG 2100  
Db  
2041 AAATTTGCAATTTGAAAAATCAACGCGAAAATTAGGAGGTATTAAACGAGAGCTTTGACTGCTCAG 2100  
QY  
2101 AAATTTGCAGAAATATCACACAGAAAGAAAGAACGGAACAAATGCGCAATTAATATGT 2160  
Db  
2101 AAATTTGCAGAAATATCACACAGAAAGAAAGAACGGAACAAATGCGCAATTAATATGT 2160  
QY  
2161 ATGTTTGGAAATTTGATGTTCACTCAACCTCTACAGTGGAAATTTGATTTCTTATAGCGG 2220  
Db  
2161 ATGTTTGGAAATTTGATGTTCACTCAACCTCTACAGTGGAAATTTGATTTCTTATAGCGG 2220  
QY  
2221 CTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGTGACTCAAG 2280  
Db  
2221 CTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGATTTGTGACTCAAG 2280  
QY  
2281 AAGAAATGCTGCCGCTGAGCGTGCAGTGGCTCATGAGCGGGAAGAAACAGATATTTTGG 2340  
Db  
2281 AAGAAATGCTGCCGCTGAGCGTGCAGTGGCTCATGAGCGGGAAGAAACAGATATTTTGG 2340  
QY  
2341 AAGCAAAAGTTTGGTGAATTTGCTCAGAGAAATTCGAGAAACAAACGACAACTCCAGCACCCAG 2400  
Db  
2341 AAGCAAAAGTTTGGTGAATTTGCTCAGAGAAATTCGAGAAACAAACGACAACTCCAGCACCCAG 2400  
QY  
2401 CGCATATTTGTAGTCTATCGAGACGGAGTTAGCGATTCGGAGATGCTACGTTGTAGTCATG 2460  
Db  
2401 CGCATATTTGTAGTCTATCGAGACGGAGTTAGCGATTCGGAGATGCTACGTTGTAGTCATG 2460  
QY  
2461 ATGAGCTTCGATCTTTTAAAAAGCGAAGTAAAAACAATTCATGTCCGAAACGCGGATGGAGAAG 2520  
Db  
2461 ATGAGCTTCGATCTTTTAAAAAGCGAAGTAAAAACAATTCATGTCCGAAACGCGGATGGAGAAG 2520  
QY  
2521 ATCCAGAGCCGAAGTACACGTTTCATTTGATTCAGAAAAAGACAAATACACGATTCCTTTC 2580  
Db  
2521 ATCCAGAGCCGAAGTACACGTTTCATTTGATTCAGAAAAAGACAAATACACGATTCCTTTC 2580  
QY  
2581 GAAGAATGGAAAAAGATAAGCCAGTGGTCAATTAAGATCTTACTCTCTGCTGAAAAACAGATG 2640  
Db  
2581 GAAGAATGGAAAAAGATAAGCCAGTGGTCAATTAAGATCTTACTCTCTGCTGAAAAACAGATG 2640  
QY  
2641 TCGCTGTTGCTGCTGTTTAAAAAATGGAGGAGGATATGAAAGAAAGCAAAAGAACTGGAA 2700  
Db  
2641 TCGCTGTTGCTGCTGTTTAAAAAATGGAGGAGGATATGAAAGAAAGCAAAAGAACTGGAA 2700  
QY  
2701 TTGTGAAACCCATCATCCGGAACAACTGTGGAATAAACTTATCGTTTGGAAATACAAATTCG 2760  
Db  
2701 TTGTGAAACCCATCATCCGGAACAACTGTGGAATAAACTTATCGTTTGGAAATACAAATTCG 2760  
QY  
2761 ATTTTCTTGTGCACTCTCATGATGTTGGTGTCTTTGGTCAATCTCGTCCAGGACATTAACATG 2820  
Db  
2761 ATTTTCTTGTGCACTCTCATGATGTTGGTGTCTTTGGTCAATCTCGTCCAGGACATTAACATG 2820  
QY  
2821 TTATGTATGACGATTAAGGAATGAGCCAGATGAAGTCTTAATAATGACTACGACTTGG 2880  
Db  
2821 TTATGTATGACGATTAAGGAATGAGCCAGATGAAGTCTTAATAATGACTACGACTTGG 2880

```
QY 2881 CTTTCTCTCTGTAGATGTCGAAACCCATCTCGTTCCTGTTCCGGTTCATTATGCTC 2940
Db 2881 CTTTCTCTCTGTAGATGTCGAAACCCATCTCGTTCCTGTTCCGGTTCATTATGCTC 2940
QY 2941 ATTATCATGTGAAAGCGAAGAGCTTTATCGAACTTACAGGAACATTACATCGGTG 3000
Db 2941 ATTATCATGTGAAAGCGAAGAGCTTTATCGAACTTACAGGAACATTACATCGGTG 3000
QY 3001 ACTATGACAGCCACGGACTCGACAGAAATGGAACTTTTCTCCTCAAACTAAAGT 3060
Db 3001 ACTATGACAGCCACGGACTCGACAGAAATGGAACTTTTCTCCTCAAACTAAAGT 3060
QY 3061 ACCCTGAAATGCTGCGATACATTTTTCGAAAGTGTCCCGGTTTCAATCAATTTT 3120
Db 3061 ACCCTGAAATGCTGCGATACATTTTTCGAAAGTGTCCCGGTTTCAATCAATTTT 3120
QY 3121 TCAATGTAGATATTGTACTTACTTTTAAAGCCCGGTTTCAAAATTCATTCATG 3180
Db 3121 TCAATGTAGATATTGTACTTACTTTTAAAGCCCGGTTTCAAAATTCATTCATG 3180
QY 3181 ACTAAGCTTTTCATAAATTAATTGAAATTTAAAGGAGGAGGAGGAGGAGGAGGAG 3227
Db 3181 ACTAAGCTTTTCATAAATTAATTGAAATTTAAAGGAGGAGGAGGAGGAGGAGGAG 3227

RESULT 3
US-11-144-985-2
; Sequence 2, Application US/11144985
; Publication No. US20060024798A1
; GENERAL INFORMATION:
; APPLICANT: Mello, Craig C.
; APPLICANT: Tabara, Hiroaki
; APPLICANT: Grishok, Alla
; APPLICANT: Fire, Andrew
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC
; FILE OF INVENTION: INTERFERENCE
; FILE REFERENCE: UMY-052CN
; CURRENT APPLICATION NUMBER: US/11/144,985
; CURRENT FILING DATE: 2005-06-03
; PRIOR APPLICATION NUMBER: US 09/689,992
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3227
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21)...(3080)
US-11-144-985-2

Query Match 100.0%; Score 3227; DB 16; Length 3227;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCCAAAGTGATGAACATGCTCGAATTTCCGAAATGGAAAGGATTTTATC 60
Db 1 CAGCCCAAAGTGATGAACATGCTCGAATTTCCGAAATGGAAAGGATTTTATC 60
QY 61 GTCAATCTCTCGATCCGAGATGAATGCTTTCGAGGCCCACTGCTTAATGCGCGCA 120
Db 61 GTCAATCTCTCGATCCGAGATGAATGCTTTCGAGGCCCACTGCTTAATGCGCGCA 120
QY 121 AATCTATGAGAAGAAAGTACTTCTTTTGGTAAATGGTTCAAGTTCTCCAGCAAAATTT 180
Db 121 AATCTATGAGAAGAAAGTACTTCTTTTGGTAAATGGTTCAAGTTCTCCAGCAAAATTT 180
```

```
QY 181 ACGATCGGGAATACTACGAGTATGAAGTAAATGACAAAGGAAGTATTGAAATAGAAAAC 240
Db 181 ACGATCGGGAATACTACGAGTATGAAGTAAATGACAAAGGAAGTATTGAAATAGAAAAC 240
QY 241 CAGGAAACCTTTCCCAAAAAGACAGAAATTCCTCAATTCCTCGATCGTGCAGAACTCTTCT 300
Db 241 CAGGAAACCTTTCCCAAAAAGACAGAAATTCCTCAATTCCTCGATCGTGCAGAACTCTTCT 300
QY 301 GGCAACATCTTTCGCGATGAGAAGACAGACAGATTTTATTTCTCGAAGACTATGTTTTTG 360
Db 301 GGCAACATCTTTCGCGATGAGAAGACAGACAGATTTTATTTCTCGAAGACTATGTTTTTG 360
QY 361 ATGAAAAGGACACTGTTTATAGTGTTCGACGTGAACACTGTGCACATCAAAAATGCTGG 420
Db 361 ATGAAAAGGACACTGTTTATAGTGTTCGACGTGAACACTGTGCACATCAAAAATGCTGG 420
QY 421 TTTTCGAGAAAGTAGTAAAGGATTTTCGAGAAAAGGATGAAAGGATTTTCGAGAAA 480
Db 421 TTTTCGAGAAAGTAGTAAAGGATTTTCGAGAAAAGGATGAAAGGATTTTCGAGAAA 480
QY 481 AAATCTTTATACAAATGATACTTACCTTATCGTAAAGGATTTTACCTGAACTTTAGTGCAG 540
Db 481 AAATCTTTATACAAATGATACTTACCTTATCGTAAAGGATTTTACCTGAACTTTAGTGCAG 540
QY 541 AAATCCGAAAAGACGAAAGCGAATCGGAGTTTACAAATTCCTGAAAGATTTATGA 600
Db 541 AAATCCGAAAAGACGAAAGCGAATCGGAGTTTACAAATTCCTGAAAGATTTATGA 600
QY 601 CCCAGAAAGTTTCCTACGCGCTTTTGTGAACAGGAGATTTAAGTACAAATTCGCGAAA 660
Db 601 CCCAGAAAGTTTCCTACGCGCTTTTGTGAACAGGAGATTTAAGTACAAATTCGCGAAA 660
QY 661 ATTTTGTGTACGATAAATTTCAATTTCTCGAGTTCCTGAATCGTTTTCAGCATCCAAACA 720
Db 661 ATTTTGTGTACGATAAATTTCAATTTCTCGAGTTCCTGAATCGTTTTCAGCATCCAAACA 720
QY 721 GATTGCAACATCAATTTAGAAAGTAGCAACGAAGATCGAAGATGGTTTGAATTTACATG 780
Db 721 GATTGCAACATCAATTTAGAAAGTAGCAACGAAGATCGAAGATGGTTTGAATTTACATG 780
QY 781 GAATCAAGAAATTTGTTTCGATGTTGAACCTGTGCTCAATTTTGCATTTTCGATAAATCT 840
Db 781 GAATCAAGAAATTTGTTTCGATGTTGAACCTGTGCTCAATTTTGCATTTTCGATAAATCT 840
QY 841 TCTACAAATGCAACGAAATTTCTCTTCTGATTTATCTTCTCTAAATTTGTCGACCCCACT 900
Db 841 TCTACAAATGCAACGAAATTTCTCTTCTGATTTATCTTCTCTAAATTTGTCGACCCCACT 900
QY 901 CGTGTAAAGATGATGACGAAAGATCTTAAACAAACCTGATGCGGCGGAAATGACAA 960
Db 901 CGTGTAAAGATGATGACGAAAGATCTTAAACAAACCTGATGCGGCGGAAATGACAA 960
QY 961 TCAGACAAGCCGCGCGCAAGATTTTCGACAAATTTTGGAAATTTTGAAGTCTGAAATGCG 1020
Db 961 TCAGACAAGCCGCGCGCAAGATTTTCGACAAATTTTGGAAATTTTGAAGTCTGAAATGCG 1020
QY 1021 CAGAGTTTGGGATTAACGAAATGTCGAGATTTGACAGAGGACATCTGACATTTCTAGATT 1080
Db 1021 CAGAGTTTGGGATTAACGAAATGTCGAGATTTGACAGAGGACATCTGACATTTCTAGATT 1080
QY 1081 TGTGCGAGGAAATCTCTCTTTGTTTATAAGTCACTGTTTAAATTCGAGGAGGAAATG 1140
Db 1081 TGTGCGAGGAAATCTCTCTTTGTTTATAAGTCACTGTTTAAATTCGAGGAGGAAATG 1140
QY 1141 CAAAAGATGACGATCTACTATGTTTCAAAATCTATGAGGAAACAAAGATTTCTATGAGT 1200
Db 1141 CAAAAGATGACGATCTACTATGTTTCAAAATCTATGAGGAAACAAAGATTTCTATGAGT 1200
QY 1201 TTTCCCACTTACCACTAGTCAAAATTTTAAAGTGGAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 1201 TTTCCCACTTACCACTAGTCAAAATTTTAAAGTGGAGGAGGAGGAGGAGGAGGAGGAG 1260
QY 1261 AACATCTTGAAGTTTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
```



1261 AACATCTTGAAGTTCATGAGAGCCCAAGATACAGAAATCGAATTCGATCTGGTATGC 1320  
1321 AAGACAAAGTTTCTAAAGCGAGCTACACGAAACCTCACGACTACAAAGAAAATACCCCTAA 1380  
1321 AAGACAAAGTTTCTAAAGCGAGCTACACGAAACCTCACGACTACAAAGAAAATACCCCTAA 1380  
1381 AATGCTGAAAGAAATGGATTTCTCTCTGAGAGCTAAATTTGTTGAAAGATTTGGAT 1440  
1381 AATGCTGAAAGAAATGGATTTCTCTCTGAGAGCTAAATTTGTTGAAAGATTTGGAT 1440  
1441 TATGCTCCAAATCTCAGATGATCGAATGTCAGGAAAGGTTTGAAGAGCCCAATGCTTG 1500  
1441 TATGCTCCAAATCTCAGATGATCGAATGTCAGGAAAGGTTTGAAGAGCCCAATGCTTG 1500  
1501 TGAATAGTGAATGAACAAATTAATAACACCAAGTGAATTCGTGGATTTCAAGAAAAC 1560  
1501 TGAATAGTGAATGAACAAATTAATAACACCAAGTGAATTCGTGGATTTCAAGAAAAC 1560  
1561 AATTGAATGTTCCCGAAAAGAACTTTGCTGCTGCTGTTTGTAGTCAACGAAACAG 1620  
1561 AATTGAATGTTCCCGAAAAGAACTTTGCTGCTGCTGTTTGTAGTCAACGAAACAG 1620  
1621 CGGAAATCCATGCTTAGAAGAGAACGACGTTGTTAAGTTCACACCGAACTAAATTTGGTG 1680  
1621 CGGAAATCCATGCTTAGAAGAGAACGACGTTGTTAAGTTCACACCGAACTAAATTTGGTG 1680  
1681 GTTGCAAGTTCGCGTGAATA CGAATTTGGTGCCAAATGAAGAACAGAGAGCGCAATCTATTA 1740  
1681 GTTGCAAGTTCGCGTGAATA CGAATTTGGTGCCAAATGAAGAACAGAGAGCGCAATCTATTA 1740  
1741 TGTAACGCGGACGAAATGAATATGCTTCTACAAATTTGTACACTTAATACCGGAA 1800  
1741 TGTAACGCGGACGAAATGAATATGCTTCTACAAATTTGTACACTTAATACCGGAA 1800  
1801 TCGGTAGATTGAATAGCGCAACAGAGCGAAGATATGTTTGAACGCTTTCCCGATA 1860  
1801 TCGGTAGATTGAATAGCGCAACAGAGCGAAGATATGTTTGAACGCTTTCCCGATA 1860  
1861 AAGAAACAAAGTCTTAATGTTCAATTCATTTCCAAACGACAACTGAATGCTTACGGTT 1920  
1861 AAGAAACAAAGTCTTAATGTTCAATTCATTTCCAAACGACAACTGAATGCTTACGGTT 1920  
1921 TTGTGAAACATTTATTCGATCACACCATCGGTGTAGCTAATCAGCATATTACTTCTGAAA 1980  
1921 TTGTGAAACATTTATTCGATCACACCATCGGTGTAGCTAATCAGCATATTACTTCTGAAA 1980  
1981 CAGTCACAAAGCTTTGGCATCACTAAGGACAGAGAAAGGATCAAAACGAAATTTTCTATC 2040  
1981 CAGTCACAAAGCTTTGGCATCACTAAGGACAGAGAAAGGATCAAAACGAAATTTTCTATC 2040  
2041 AAATTCATTTGAAATCAACCGGAAATAGGAGTTAATTAACAGGAGCTTGACTGTCAG 2100  
2041 AAATTCATTTGAAATCAACCGGAAATAGGAGTTAATTAACAGGAGCTTGACTGTCAG 2100  
2101 AAATTCAGAAATATCACAGAGAAAGAAAGAGCGGAAACAAATGCCATTAATCTATGT 2160  
2101 AAATTCAGAAATATCACAGAGAAAGAAAGAGCGGAAACAAATGCCATTAATCTATGT 2160  
2161 ATGTTGGAATTTGATGTAACTCATCAACCTCTACAGTGAATTTGATTTATTTATAGCGG 2220  
2161 ATGTTGGAATTTGATGTAACTCATCAACCTCTACAGTGAATTTGATTTATTTATAGCGG 2220  
2221 CTGTAGTACGAGTATCAATCCAGGTGGAACCTATCTATCGAAATATGATTTGACTCAAG 2280  
2221 CTGTAGTACGAGTATCAATCCAGGTGGAACCTATCTATCGAAATATGATTTGACTCAAG 2280  
2281 AAGAATGTCCTCCCGTAGCGTGCAGTGCCTCATGACGGAAGAAACAGATATTTTCG 2340  
2281 AAGAATGTCCTCCCGTAGCGTGCAGTGCCTCATGACGGAAGAAACAGATATTTTCG 2340  
2341 AAGCAAAGTTCGTGAAATTTGCTCAGAGAAATTCGAGAAACCAACGACCAATCGAGCACACG 2400

2341 AAGCAAAGTTCGTGAAATTTGCTCAGAGAAATTCGAGAAACCAACGACCAATCGAGCACACG 2400  
2401 CGCATATTTGTAGTCTATCGAGCGGAGTTAGCGATTCGAGAGATGCTACGTGTAGTCAATG 2460  
2401 CGCATATTTGTAGTCTATCGAGCGGAGTTAGCGATTCGAGAGATGCTACGTGTAGTCAATG 2460  
2461 ATGAGCTTCGATCTTTTAAAGAGCGAAGTAAACAAATTCATGTCGGAACGGATGGAGAAG 2520  
2461 ATGAGCTTCGATCTTTTAAAGAGCGAAGTAAACAAATTCATGTCGGAACGGATGGAGAAG 2520  
2521 ATCCAGAGCCGAGTACACGTTTCAATTTGATTCAGAAAAGACACAATACACGATTCCTTC 2580  
2521 ATCCAGAGCCGAGTACACGTTTCAATTTGATTCAGAAAAGACACAATACACGATTCCTTC 2580  
2581 GAAGAATGGAAGAAAGATTAAGCCAGTGGTCAATTAAGATCTTACTCTGCTGTAACACAGATG 2640  
2581 GAAGAATGGAAGAAAGATTAAGCCAGTGGTCAATTAAGATCTTACTCTGCTGTAACACAGATG 2640  
2641 TCGCTGTTGCTGCTGTTAAACAAATCGGAGGAGGATATGAAGAAAGCAAGAACTGGAA 2700  
2641 TCGCTGTTGCTGCTGTTAAACAAATCGGAGGAGGATATGAAGAAAGCAAGAACTGGAA 2700  
2701 TTGTGAACCCCATCATCCGAAACAACTGGATTAATCTTATCTGTTTTCGAAATACAAATTCG 2760  
2701 TTGTGAACCCCATCATCCGAAACAACTGGATTAATCTTATCTGTTTTCGAAATACAAATTCG 2760  
2761 ATTTTCTTGGCATCTCATGCTGCTTGGTACATCTCTGTCGACGACATTAACATG 2820  
2761 ATTTTCTTGGCATCTCATGCTGCTTGGTACATCTCTGTCGACGACATTAACATG 2820  
2821 TTATGCTATGACGATTAAGGAATGAGCCAGAGTGAAGTCTTATTAATGACTACGAGACTTG 2880  
2821 TTATGCTATGACGATTAAGGAATGAGCCAGAGTGAAGTCTTATTAATGACTACGAGACTTG 2880  
2881 CTTTTCTCTCTGCTAGATGT CGAAAACCCATCTCGTTGCTGTTCCGGTTTCAATTCGCTC 2940  
2881 CTTTTCTCTCTGCTAGATGT CGAAAACCCATCTCGTTGCTGTTCCGGTTTCAATTCGCTC 2940  
2941 ATTTATCATGTGAAAAGGAAAGAGCTTTATCGAATCTTACAGGAACATTAACATCGGTG 3000  
2941 ATTTATCATGTGAAAAGGAAAGAGCTTTATCGAATCTTACAGGAACATTAACATCGGTG 3000  
3001 ACTATGACAGCCAGGACTCGACACGAAATCGAAATCGAAATCTTCCAAACTAACGCTGAAAT 3060  
3001 ACTATGACAGCCAGGACTCGACACGAAATCGAAATCGAAATCTTCCAAACTAACGCTGAAAT 3060  
3061 ACCCTGGAATGCTGTTGCGATTAACATTTTGAAGAGTGTGCGCCGTTTCAATCAAAATTTT 3120  
3061 ACCCTGGAATGCTGTTGCGATTAACATTTTGAAGAGTGTGCGCCGTTTCAATCAAAATTTT 3120  
3121 TCAATTTGATGATTTGATTTACTTTTAAAGCCCGGTTTCAAAAATTCATTCATCCATG 3180  
3121 TCAATTTGATGATTTGATTTACTTTTAAAGCCCGGTTTCAAAAATTCATTCATCCATG 3180  
3181 ACTAACGTTTTCATTAATTTACTTTGAAATTTTAAAGAAAAA 3227  
3181 ACTAACGTTTTCATTAATTTACTTTGAAATTTTAAAGAAAAA 3227

RESULT 4  
US-10-645-746-1  
; Sequence 1, Application US/10645746  
; Publication No. US20040265839A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; FILE REFERENCE: UNY-052DV1  
; CURRENT APPLICATION NUMBER: US/10/645,746  
; CURRENT FILING DATE: 2003-08-20

```

; PRIOR APPLICATION NUMBER: US 09/689,992
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3719
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-10-645-746-1

```

Query Match	80.6%	Score 2601;	DB 9;	Length 3719;
Best Local Similarity	86.3%	Pred. No. 0;		
Matches 3210;	Conservative 0;	Mismatches 0;	Indels 509;	Gaps 10;
QY 1	CAGCCACAAAGTGTATGAACAATGTCCTCGAATTTTCCGGAATTCGAAAAAGGATTTTATC	60		
DB 1				
QY 61	GTCAATCTCTCGATCC-----	76		
DB 61	GTCAATCTCTCGATCCGGTATGATCAATTTATGACGCTATAAGATATATAAGTTTGTGATA	120		
QY 77	-----GGAGATGAATGCTTTCGAGGCCCACTGCTGAATATCGACGGCAAAATTC	125		
DB 121	TTAATATTAATAGGAGATGAATGCTTTCGAGGCCCACTGGTAAATCGGACGGCAAAATTC	180		
QY 126	TATGAGAAGAAAGTACTCTTTTGGTGAATTTGGTTCAAGTTCTCCAGCAAAATTTACGAT	185		
DB 181	TATGAGAAGAAAGTACTCTTTTGGTGAATTTGGTTCAAGTTCTCCAGCAAAATTTACGAT	240		
QY 186	CGGGAATACTACGAGTATGAAGTGAAATTCGACAAAGGAAGTATTGTAATGAAACCCAGGA	245		
DB 241	CGGGAATACTACGAGTATGAAGTGAAATTCGACAAAGGAAGTATTGTAATGAAACCCAGGA	300		
QY 246	AAACCTTTCCCAAAAAGACAGAAATTCCAAT-----	277		
DB 301	AAACCTTTCCCAAAAAGACAGAAATTCCAATGTAAGTGCCTGTGAATTAGTCAAAACTA	360		
QY 278	-----TCCCGATCGTGCAAAACCTCTTCTTGCGCAACATCTTCGGCATGAGAAG	323		
DB 361	ATTTTATTTTTCAGTCCCGATCGTGCAAAACCTCTTCTTGCGCAACATCTTCGGCATGAGAAG	420		
QY 324	AAGCACACAGATTTTATCTCGAGACTATGTTTTGATGAAAAGGACACTGTTTATAGT	383		
DB 421	AAGCACACAGATTTTATCTCGAAGACTATGTTTTGATGAAAAGGACACTGTTTATAGT	480		
QY 384	GTTTGTGCACTGAACACTGTGCACATCAAAAATGCTGTTTTCGGAGAAAGTAGTAAAAAAG	443		
DB 481	GTTTGTGCACTGAACACTGTGCACATCAAAAATGCTGTTTTCGGAGAAAGTAGTAAAAAAG	540		
QY 444	GATTCGGAGAAAAAGATGAAAAAGGATTCGGAGAAAAAAATCTTTATACAAATGATACCTT	503		
DB 541	GATTCGGAGAAAAAGATGAAAAAGGATTCGGAGAAAAAAATCTTTATACAAATGATACCTT	600		
QY 504	ACCTATCGTAAAAAATTTACCTGGAACCTTTAGTTCGAGAAAAATCGGAAAAAGACGAAGAA	563		
DB 601	ACCTATCGTAAAAAATTTACCTGGAACCTTTAGTTCGAGAAAAATCGGAAAAAGACGAAGAA	660		
QY 564	GCGAATCGGAGTTACAAATTCCTGA-----	588		
DB 661	GCGAATCGGAGTTACAAATTCCTGGAAGGTTTATGAAAAACACCGATTATTAACAAAAA	720		
QY 589	-----AGAATGTTATGACCCAGAAAGTTTCGCTACGCGCCCTTTTGTGAAACGAGAGA	639		
DB 721	TTTAGCTTTTCAGAATGTTATGACCCAGAAAGTTTCGCTACGCGCCCTTTTGTGAAACGAGAGA	780		
QY 640	TTAAAGT-----	646		
DB 781	TTAAAGTGTGAGTTGCAATAATAATAATAATACCTCAACTCACTTTATATATTTTAA	840		

Qy	647	-ACAATTGCGGAAAAATTTTGTGTACGATAATAATTCAATTCTGCGAGTTCTCTGAATCGT	705
Db	841	GACAAATTCGCGAAAAATTTTGTGTACGATAATAATTCAATTCTGCGAGTTCTCTGAATCGT	900
Qy	706	TTACAGATCCAAACAGAGATTCGAAACAATCATTTAGAGTAGCACCAGAAATCGAAGCATGGT	765
Db	901	TTACAGATCCAAACAGAGATTCGAAACAATCATTTAGAGTAGCACCAGAAATCGAAGCATGGT	960
Qy	766	TTGGAAATTACATTTGGGAATCAAGAAATGTTTCGATGGTGAAACCTGTGCTCAATTTTGC	825
Db	961	TTGGAAATTACATTTGGGAATCAAGAAATGTTTCGATGGTGAAACCTGTGCTCAATTTTGC	1020
Qy	826	-----TTGTCGATAAACT	838
Db	1021	GTAAGTTTGGAAAACTCGGATAAAAAATCATGTGATTTTGTGTGAAGTTTGTGATAAACT	1080
Qy	839	ATTCTACAATGCACCGAAAAATGTCTCTTCGGATTAATCTTCTCCTTAATTTGTGCAACCCCA	898
Db	1081	ATTCTACAATGCACCGAAAAATGTCTCTTCGGATTAATCTTCTCCTTAATTTGTGCAACCCCA	1140
Qy	899	GTCGTGTAAACGATGATGTACGAAAAGATCTTAAAAACAATACTGATGCGCGGAAAAATGAC	958
Db	1141	GTCGTGTAAACGATGATGTACGAAAAGATCTTAAAAACAATACTGATGCGCGGAAAAATGAC	1200
Qy	959	AATCAGACAAGCCGCGGCAAGAAATTCGACAAATTAATTGGAATAATTTGAAGCTGAAATG	1018
Db	1201	AATCAGACAAGCCGCGGCAAGAAATTCGACAAATTAATTGGAATAATTTGAAGCTGAAATG	1260
Qy	1019	CGCAGAAGTTTGGATAACGAAAT-----	1042
Db	1261	CGCAGAAGTTTGGATAACGAAATGTTAGTTTAAATTAATTCAACAATTAATAACA	1320
Qy	1043	-----GTGAGATTCACAGAACGACATCTGACATTTCTAGATTTGTGCGAGGAAA	1092
Db	1321	TGATTTTCAGGTGAGATTGACAGACGACATCTGACATTTCTAGATTTGTGCGAGGAAA	1380
Qy	1093	ACTCTCTGTTTATAAAGTCACTGGTAAATTCGACAGAGAGAGAAATGCAAAAAAGTACG	1152
Db	1381	ACTCTCTGTTTATAAAGTCACTGGTAAATTCGACAGAGAGAGAAATGCAAAAAAGTACG	1440
Qy	1153	ATACTACATTTGTTCAAAATCTATGAGGAAAAAATAAAGTTTCATTGAGTTTCCCACTAC	1212
Db	1441	ATACTACATTTGTTCAAAATCTATGAGGAAAAAATAAAGTTTCATTGAGTTTCCCACTAC	1500
Qy	1213	CACGTGTCAAAGTTTAAAGTGGAGCAAAAGATACGCTCTGACCAATGGACATCTTTGAAG	1272
Db	1501	CACGTGTCAAAGTTTAAAGTGGAGCAAAAGATACGCTCTGACCAATGGACATCTTTGAAG	1560
Qy	1273	TTCAATGAGAAGCCACAAAGATACAGAAATCGAATTGATCTGGTGATGCAAGCAAGTTTC	1332
Db	1561	TTCAATGAGAAGCCACAAAGATACAGAAATCGAATTGATCTGGTGATGCAAGCAAGTTTC	1620
Qy	1333	TAAAGCGAGCTACAGAAAACTTCACGACTACAAGAAAAATACCTTAAATAATGCTGAAAG	1392
Db	1621	TAAAGCGAGCTACAGAAAACTTCACGACTACAAGAAAAATACCTTAAATAATGCTGAAAG	1680
Qy	1393	AATTTGATTTCTCTTCTGAGAGCTTAAATTTTGTGTGAAGATTTGATTTATGCTCCAAAC	1452
Db	1681	AATTTGATTTCTCTTCTGAGAGCTTAAATTTTGTGTGAAGATTTGATTTATGCTCCAAAC	1740
Qy	1453	TTCAATGATTCGAATGTCAGGAAAAGTTTTCGAAGAGCCAATGCTTTGTAATAGTGTAA	1512
Db	1741	TTCAATGATTCGAATGTCAGGAAAAGTTTTCGAAGAGCCAATGCTTTGTAATAGTGTAA	1800
Qy	1513	ATGAAACAATTTAAATGACACAGTGATTCGTGGATTTTCAAGAAAAACAATTTGAATGTGG	1572
Db	1801	ATGAAACAATTTAAATGACACAGTGATTCGTGGATTTTCAAGAAAAACAATTTGAATGTGG	1860
Qy	1573	TTCCGAAAAAGAACCTTTGCTGTGCTGTTTTTTGTAGTCAACGAAACGCGGAAATTCAT	1632
Db	1861	TTCCGAAAAAGAACCTTTGCTGTGCTGTTTTTTGTAGTCAACGAAACGCGGAAATTCAT	1920

QY 1633 GCTTAGAAGAACGAGCTTGT----- 1654  
 DB 1921 GCTTAGAGAGACGAGCTTGTGAAGTGTCTTACGTAGATATTCCGAAATATTTTC 1980  
 QY 1655 --TAAGTTCTACACCGAACTAAATGGTGGTGTGCAAGTTCGGTGAATACGAATGGTGCC 1712  
 DB 1981 AGTAAGTTCTACACCGAACTAAATGGTGGTGTGCAAGTTCGGTGAATACGAATGGTGCC 2040  
 QY 1713 AATGAANAACAGAGGAGCGCAATCTATATGTAGACGCGCAAGAAATGAATATGCC--- 1769  
 DB 2041 AATGAANAACAGAGGAGCGCAATCTATATGTAGACGCGCAAGAAATGAATATGCCGTA 2100  
 QY 1770 -----TTCTACAAAAATTTGACAC 1788  
 DB 2101 AGTTTCAGAAAAATTGAAAGTTTTTAAATATCATATTACAGTTCTACAAAAATTTGACAC 2160  
 QY 1789 TAAATACCGGAATCGGTAGATTGAAATAGCCGCAACAGAAGCGAAGAATATGTTTGAAC 1848  
 DB 2161 TAAATACCGGAATCGGTAGATTGAAATAGCCGCAACAGAAGCGAAGAATATGTTTGAAC 2220  
 QY 1849 GTCTTCCGATAAAGAACAAAAAGTCTTAAATGTTTCAATTCATCTTCCAAACGCAACTGA 1908  
 DB 2221 GTCTTCCGATAAAGAACAAAAAGTCTTAAATGTTTCAATTCATCTTCCAAACGCAACTGA 2280  
 QY 1909 ATGCTTACCGTCTTGTGAACATTAATGCGATCACACCATCGGTAGCTAATTCAGCATA 1968  
 DB 2281 ATGCTTACCGTCTTGTGAACATTAATGCGATCACACCATCGGTAGCTAATTCAGCATA 2340  
 QY 1969 TTACTTCTGAAACAGTCAACAAAGCTTTGGCATCACTAAGGCACGAGAAAGGATCAAAAC 2028  
 DB 2341 TTACTTCTGAAACAGTCAACAAAGCTTTGGCATCACTAAGGCACGAGAAAGGATCAAAAC 2400  
 QY 2029 GAATTTTCTATCAAAATTCGAAATCAACGCGAAATTAGAGGTATTAACGAGGAC 2088  
 DB 2401 GAATTTTCTATCAAAATTCGAAATCAACGCGAAATTAGAGGTATTAACGAGGAC 2460  
 QY 2089 TTGACTGTGTCAGAAATTCGAGAAATATCACAGAGAAAGAAAGACGCGAAACAAATGC 2148  
 DB 2461 TTGACTGTGTCAGAAATTCGAGAAATATCACAGAGAAAGAAAGACGCGAAACAAATGC 2520  
 QY 2149 CATTAACATATGATGTGGAATTTGATGTAACTCATCCAACTCTCACTAGTGGAAATTTGATT 2208  
 DB 2521 CATTAACATATGATGTGGAATTTGATGTAACTCATCCAACTCTCACTAGTGGAAATTTGATT 2580  
 QY 2209 ATCTATACCGCTGTAGTAGCAGATATCAATCCAGGTGGAATCTATCCAAATATGA 2268  
 DB 2581 ATCTATACCGCTGTAGTAGCAGATATCAATCCAGGTGGAATCTATCCAAATATGA 2640  
 QY 2269 TTGTGACTCAAGAAATTCGTCGCGGTGAGCGTGCAGTGGCTCATGGACGCGAAAGAA 2328  
 DB 2641 TTGTGACTCAAGAAATTCGTCGCGGTGAGCGTGCAGTGGCTCATGGACGCGAAAGAA 2700  
 QY 2329 CAGATATTTTGGAGCAAAAGTTGCGTGAATTTGCTCAGAGAAATTCGCAGAA----- 2378  
 DB 2701 CAGATATTTTGGAGCAAAAGTTGCGTGAATTTGCTCAGAGAAATTCGCAGAGAAATTCGCAGAA 2760  
 QY 2379 -----AACACGAC 2387  
 DB 2761 TTGAGTATTTAAAGATCTCTGGATTTTTTAAATTTTTTTTGTAAACTTTTCAGAACACGAC 2820  
 QY 2388 AATCGAGCACCGCGCATATTTAGTCTATCGAGACGGAGTTAGCGATTCGAGATGCTA 2447  
 DB 2821 AATCGAGCACCGCGCATATTTAGTCTATCGAGACGGAGTTAGCGATTCGAGATGCTA 2880  
 QY 2448 CCGTGTAGTCATGATGAGCTTCGATCTTTTAAAGCGAAAGTAAACAAATTCATGTCGAA 2507  
 DB 2881 CCGTGTAGTCATGATGAGCTTCGATCTTTTAAAGCGAAAGTAAACAAATTCATGTCGAA 2940  
 QY 2508 CCGGATGGAGAGATCCAGCGCGAGTACAGTTTCATTTGATTCAGAAAGACACAT 2567  
 DB 2941 CCGGATGGAGAGATCCAGCGCGAGTACAGTTTCATTTGATTCAGAAAGACACAT 3000  
 QY 2568 ACACGATTCTTCGAAGAATGGAAAAAGATAAGCCAGTGGTCAATAAAGATCTTACTCCT 2627

DB 3001 ACACGATTCGTTTCGAAGATGGAAAAAGATAGCCAGTGGTCAATAAAGATCTTACTCCT 3060  
 QY 2628 GCTGAAACAGATGTCGCTGTTCTGCTGTGTAAACAATGGGAGGAGATATGAAGAAAGC 2687  
 DB 3061 GCTGAAACAGATGTCGCTGTTCTGCTGTGTAAACAATGGGAGGAGATATGAAGAAAGC 3120  
 QY 2688 AAGAAACCTGGAAATTTGTGAACCCATCATCCGAAACAACCTGTGGATAAACTTATCGTTTCG 2747  
 DB 3121 AAGAAACCTGGAAATTTGTGAACCCATCATCCGAAACAACCTGTGGATAAACTTATCGTTTCG 3180  
 QY 2748 AATACAAATTCGAATTTTTTCTTGGCATCTCATCTATGGTGTCTTGGTATCATCTCGTCCA 2807  
 DB 3181 AATACAAATTCGAATTTTTTCTTGGCATCTCATCTATGGTGTCTTGGTATCATCTCGTCCA 3240  
 QY 2808 GGACATTCACCTGTTATGTATGACGATAAAGGAATGAGCAAGATGAAGTCTAT 2861  
 DB 3241 GGACATTCACCTGTTATGTATGACGATAAAGGAATGAGCAAGATGAAGTCTATGTAAGC 3300  
 QY 2862 ----- 2861  
 DB 3301 GTTTTGAATAGCAGTTAGCGATTTTAGGATTTTGTAAATCCGCATATAGTTATTTATAAAA 3360  
 QY 2862 -----AAAAATGACCTACGGACTTGTCTTTTCTCTCTGTAGATGTGAAAAACCCAT 2911  
 DB 3361 AATGTTTTCAGAAAAATGACCTACGGACTTGTCTTTTCTCTCTGTAGATGTGAAAAACCCAT 3420  
 QY 2912 CTCGTTGCTGTTCCGGTTCATATGCTCATTTATCATGTGAAAAAGCGAAAGAGCTTTA 2971  
 DB 3421 CTCGTTGCTGTTCCGGTTCATATGCTCATTTATCATGTGAAAAAGCGAAAGAGCTTTA 3480  
 QY 2972 TCGAACTTACAGGAACATTAACATCGGTGACTATGCACAGCCACGAGCTCGACACGAAAT 3031  
 DB 3481 TCGAACTTACAGGAACATTAACATCGGTGACTATGCACAGCCACGAGCTCGACACGAAAT 3540  
 QY 3032 GGAACATTTTCTCCAAACTAACGCTGAAAGTACCTCGAATGTCGTTTCGCATAACATTTTCG 3091  
 DB 3541 GGAACATTTTCTCCAAACTAACGCTGAAAGTACCTCGAATGTCGTTTCGCATAACATTTTCG 3600  
 QY 3092 AAAAGTGTGCGCGTTTCAATCAAAATTTTCAATTTGTAGATATTTTACTTTT 3151  
 DB 3601 AAAAGTGTGCGCGTTTCAATCAAAATTTTCAATTTGTAGATATTTTACTTTT 3660  
 QY 3152 AAAAGTGTGCGCGTTTCAAAATTTTCAATTCATTCGACTAAAGTTCATATAAATTTTGAATTT 3210  
 DB 3661 AAAAGTGTGCGCGTTTCAAAATTTTCAATTCATTCGACTAAAGTTCATATAAATTTTGAATTT 3719

RESULT 5

US-10-645-735-1  
 ; Sequence 1, Application US/10645735  
 ; Publication No. US20050100913A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mello, Craig C.  
 ; APPLICANT: Tabara, Hiroaki  
 ; APPLICANT: Grishok, Alla  
 ; APPLICANT: Fire, Andrew  
 ; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
 ; FILE REFERENCE: INTERFERENCE  
 ; CURRENT APPLICATION NUMBER: US/10/645,735  
 ; PRIOR FILING DATE: 2003-08-20  
 ; PRIOR APPLICATION NUMBER: US/09/689,992A  
 ; PRIOR FILING DATE: 2000-10-13  
 ; PRIOR APPLICATION NUMBER: US 60/193,218  
 ; PRIOR FILING DATE: 2000-03-30  
 ; PRIOR APPLICATION NUMBER: US 60/159,776  
 ; PRIOR FILING DATE: 1999-10-15  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 3719  
 ; TYPE: DNA

i: ORGANISM: Caenorhabditis elegans  
US-10-645-735-1

Query Match 80.6%; Score 2601; DB 10; Length 3719;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 3210; Conservative 0; Mismatches 0; Indels 509; Gaps 10;

QY 1 CAGCCAGAGTGTGAAACATGCTCCCGAATTTCCCGAATTCGGAAGGATTTATC 60  
DB 1 CAGCCAGAGTGTGAAACATGCTCCCGAATTTCCCGAATTCGGAAGGATTTATC 60

QY 61 GTCATTCTCTCGATCC----- 76  
DB 61 GTCATTCTCTCGATCCGATGATGATCAATTTATGACGCTATAGATATATAAGTTTGATA 120

QY 77 -----GGAGATGAAATGCTTTGCGAGGCCCACTGGTAAATCGCGCAAAATTC 125  
DB 121 TTAATATTAAGAGATGAAATGCTTTGCGAGGCCCACTGGTAAATCGCGCAAAATTC 180

QY 126 TATGAGAGAAAGTACTCTCTTTGGTAAATTTGGTCAAGTTCTCCAGCAAAATTTACGAT 185  
DB 181 TATGAGAGAAAGTACTCTCTTTGGTAAATTTGGTCAAGTTCTCCAGCAAAATTTACGAT 240

QY 186 CGGGAATACTACGAGTATGAAGTAAATGCAAAAGGAAGTATTTGAATAGAAACACGAGA 245  
DB 241 CGGGAATACTACGAGTATGAAGTAAATGCAAAAGGAAGTATTTGAATAGAAACACGAGA 300

QY 246 AAACCTTTCCCAAAAGACAGAAATTTCCAA----- 277  
DB 301 AAACCTTTCCCAAAAGACAGAAATTTCCAAATGTAAGTGTCTGTAAATTAGTCAAACTA 360

QY 278 -----TCCGATCGTGCAAACTCTCTTGCGCAACATCTTCGGCATGAGAAG 323  
DB 361 ATTTTATTTTTCAGTCCCGATCGTGCAAACTCTCTTGCGCAACATCTTCGGCATGAGAAG 420

QY 324 AAGCAGACAGATTTTATCTCGAAGACTATGTTTTGTGATGAAAGGACACTGTTTATAGT 383  
DB 421 AAGCAGACAGATTTTATCTCGAAGACTATGTTTTGTGATGAAAGGACACTGTTTATAGT 480

QY 384 GTTTGTGCGACTGAACACTGTGCATCAAAATGCTGTTTTCGGAGAAAGTAGTAAAAAAG 443  
DB 481 GTTTGTGCGACTGAACACTGTGCATCAAAATGCTGTTTTCGGAGAAAGTAGTAAAAAAG 540

QY 444 GATTCGGAGAAAAGATGAAAAGGATTTGAGAAAAAATCTTTATACAAATGATCTT 503  
DB 541 GATTCGGAGAAAAGATGAAAAGGATTTGAGAAAAAATCTTTATACAAATGATCTT 600

QY 504 ACCTATCGTAAAAATTTTCACTGAACTTTTGTAGTCGAGAAATCCGGAAGACGAGAA 563  
DB 601 ACCTATCGTAAAAATTTTCACTGAACTTTTGTAGTCGAGAAATCCGGAAGACGAGAA 660

QY 564 GCGAATCGGAGTTACAAATTCCTGA----- 588  
DB 661 GCGAATCGGAGTTACAAATTCCTGAAGGTTTATGAAAAACACGCAATTATACAAACAAA 720

QY 589 -----AGAATGTTATGACCCAGAAAGTTTCGCTACGCGCTTTTGTGAAACGAGAGA 639  
DB 721 TTAGCTTTTCAGAAATGTTATGACCCAGAAAGTTTCGCTACGCGCTTTTGTGAAACGAGAGA 780

QY 640 TTAAGT----- 646  
DB 781 TTAAGTGTGAGTTGCAATTAATTAATAATATCACTCACTCAATTTATATATTTTAA 840

QY 647 -ACAAATCGGAAAAATTTTGTAGCAATAAATTTCAATTTCTCGAGTTTCTCTGAATCGT 705  
DB 841 GACAAATCGGAAAAATTTTGTAGCAATAAATTTCAATTTCTCGAGTTTCTCTGAATCGT 900

QY 706 TTCAGATCAAAACAGATTCGAAACAATCATATTAGAAGTAGCAACAAGATCGAAGCATGGT 765  
DB 901 TTCAGATCAAAACAGATTCGAAACAATCATATTAGAAGTAGCAACAAGATCGAAGCATGGT 960

QY 766 TTGGAATTTACATTCGAATCAAGAAATTTGTCATGCTGAACTGCTCAATTTTGCAG 825  
DB 766 TTGGAATTTACATTCGAATCAAGAAATTTGTCATGCTGAACTGCTCAATTTTGCAG 825

DB 961 TTGGAATTTACATTTGGAATCAAAAGAAATTTGTTGATGTTGGAACCTGTGCTCAATTTTGCA 1020

QY 826 -----TTGTCGATAAACT 838  
DB 1021 GTAAGTTTGAGAAAACTCGGATAAAAAATCATGTGATTTTGTGGAAGTTGTGCGATAACT 1080

QY 839 ATTCTACAATGCAACCGAAATGCTCTCTTCTGGATTTATCTTCTCTAAATTTGTGCGACCCCA 898  
DB 1081 ATTCTACAATGCAACCGAAATGCTCTCTTCTGGATTTATCTTCTCTAAATTTGTGCGACCCCA 1140

QY 899 GTCGTGTAAACGATGATGTACGAAAGATCTTTAAACAAAACTGATGCGGGAATAATGAC 958  
DB 1141 GTCGTGTAAACGATGATGTACGAAAGATCTTTAAACAAAACTGATGCGGGAATAATGAC 1200

QY 959 AATCAGACAAAGCGCGCGCAAGAAATTCGACAAATTTTGAAGCTGGAATG 1018  
DB 1201 AATCAGACAAAGCGCGCGCAAGAAATTCGACAAATTTTGAAGCTGGAATG 1260

QY 1019 CGCAGAAGTTTGGGATAACGAAAT----- 1042  
DB 1261 CGCAGAAGTTTGGGATAACGAAATGTTAGTTTAAATTTATTTCAACAAATTAATATACAAAT 1320

QY 1043 -----GTGAGATTGACAGACGACATCTGACATTTCTAGATTTGTGCGAGGAAA 1092  
DB 1321 TGATTTTCAGGTGAGATTGACAGACGACATCTGACATTTCTAGATTTGTGCGAGGAAA 1380

QY 1093 ACTCTCTGTTTATAAAGTCACTGGTAAATCGGACAGAGAGAAATGCAAAAAGATGACG 1152  
DB 1381 ACTCTCTGTTTATAAAGTCACTGGTAAATCGGACAGAGAGAAATGCAAAAAGATGACG 1440

QY 1153 ATACTACATTTGTTCAAAATCTATGAGGAAAAACAAAAGTTTCATTTGAGTTTCCCACTAC 1212  
DB 1441 ATACTACATTTGTTCAAAATCTATGAGGAAAAACAAAAGTTTCATTTGAGTTTCCCACTAC 1500

QY 1213 CACTAGTCAAGTTTAAAGTGGAGCAAAAGAAATACCTGTACCAATCGAAACATCTTCAAG 1272  
DB 1501 CACTAGTCAAGTTTAAAGTGGAGCAAAAGAAATACCTGTACCAATCGAAACATCTTCAAG 1560

QY 1273 TTCATGAGAGGCCAACAAAGATACAAAGATCGAAATTTGATCTGGTGTGCAAGACAAAGTTTC 1332  
DB 1561 TTCATGAGAGGCCAACAAAGATACAAAGATCGAAATTTGATCTGGTGTGCAAGACAAAGTTTC 1620

QY 1333 TAAAGCGAGGTACAGAAACCTCACGATACAAAGAAATACCTTAAATAATGCTGAAG 1392  
DB 1621 TAAAGCGAGGTACAGAAACCTCACGATACAAAGAAATACCTTAAATAATGCTGAAG 1680

QY 1393 AATTGATTTCTCTCTGAAAGAGCTAAAATTTTCTGAAAGATTTGGAATTTGATCTCCAAAC 1452  
DB 1681 AATTGATTTCTCTCTGAAAGAGCTAAAATTTTCTGAAAGATTTGGAATTTGATCTCCAAAC 1740

QY 1453 TTCAGATGATCGAATGTCAGGAAAGGTTTGAAGAGGCCAATGCTTGTGAATAGTGTA 1512  
DB 1741 TTCAGATGATCGAATGTCAGGAAAGGTTTGAAGAGGCCAATGCTTGTGAATAGTGTA 1800

QY 1513 ATGAACAAATTAATAATGACACGAGTATTCGAGATTTCAAGAAAAACAATGGAATGTGG 1572  
DB 1801 ATGAACAAATTAATAATGACACGAGTATTCGAGATTTCAAGAAAAACAATGGAATGTGG 1860

QY 1573 TTCCCGAAAAAGAACTTTCTGCTGTTTGTAGTCAACGAAACAGCGGGAATCCAT 1632  
DB 1861 TTCCCGAAAAAGAACTTTCTGCTGTTTGTAGTCAACGAAACAGCGGGAATCCAT 1920

QY 1633 GCTTAGAAGAGAACGACGTTGT----- 1654  
DB 1921 GCTTAGAAGAGAACGACGTTGTGTAAGTGTTCAGTAGATTTTCCGAAATATTTTC 1980

QY 1655 --TAAGTCTACACCGAACTAAATTTGTTGTTGCAAGTTCGCTGGAATACGAAATGTGGCC 1712  
DB 1981 AGTAAGTTCTACACCGAACTAAATTTGTTGTTGCAAGTTCGCTGGAATACGAAATGTGGCC 2040

QY 1713 AATGAAAAACAGAGGAGCGCAATCTATTATGACGCGCAAGAAATGAATATGCGTA 1769  
DB 2041 AATGAAAAACAGAGGAGCGCAATCTATTATGACGCGCAAGAAATGAATATGCGTA 2100

Qy	1770	-----TTCTACAAAAATTGTACAC	1788
Db	2101	AGTTTTCAGAAAAATTGAAAGTTTTTAAATATCATATTTACAGTTCACAAAAATTGTACAC	2160
Qy	1789	TAAATACCGGAATCGGTAGATTGAAATAGCCGCAACAGAAGCGAAGAAATATGTTGAAAC	1848
Db	2161	TAAATACCGGAATCGGTAGATTGAAATAGCCGCAACAGAAGCGAAGAAATATGTTGAAAC	2220
Qy	1849	GTCCTCCGGATAAAGAACAAAAGCTCTTAATGTTTCATTATCATTTCCAAACGACAACCTGA	1908
Db	2221	GTCCTCCGGATAAAGAACAAAAGCTCTTAATGTTTCATTATCATTTCCAAACGACAACCTGA	2280
Qy	1909	ATGCTTACGGTTTTGTGAAACATTATTTGCATCACACCATCGGTGTAGCTTAATCAGCATTA	1968
Db	2281	ATGCTTACGGTTTTGTGAAACATTATTTGCATCACACCATCGGTGTAGCTTAATCAGCATTA	2340
Qy	1969	TTACTCTGAAACAGTCACAAAAGCTTTGGCATCATTAAGCGACGAGAAAGGATCAAAC	2028
Db	2341	TTACTCTGAAACAGTCACAAAAGCTTTGGCATCATTAAGCGACGAGAAAGGATCAAAC	2400
Qy	2029	GAAATTTCTATCAAAATTGCATTGAAATCAACGCCAAATTTAGGAGGTATTAAACCAAGAGC	2088
Db	2401	GAAATTTCTATCAAAATTGCATTGAAATCAACGCCAAATTTAGGAGGTATTAAACCAAGAGC	2460
Qy	2089	TTGACTCGGTGAGAAATTCGAAATATCACAGAGAAGAAAAGAACGCGAAACCAATGC	2148
Db	2461	TTGACTCGGTGAGAAATTCGAGNAATATCACAGAGAAGAAAAGAACGCGAAACCAATGC	2520
Qy	2149	CATTAACTATGTATGTTGGAAATTGATGTAACTCATCCAACCTCCTACAGTGGAAATTGATT	2208
Db	2521	CATTAACTATGTATGTTGGAAATTGATGTAACTCATCCAACCTCCTACAGTGGAAATTGATT	2580
Qy	2209	ATTCATAGCGGCTGTAGTAGCAGATATCAATCCAGGTGGAATCTATCTACGAAATATGA	2268
Db	2581	ATTCATAGCGGCTGTAGTAGCGNGATCAATCCAGGTGGAATCTATCTACGAAATATGA	2640
Qy	2269	TTGTGACTCAAGAAAGATGTCGTCGCCGCTGAGCGTGCAGTGGCTCATGAGACGGGAAGAA	2328
Db	2641	TTGTGACTCAAGAAAGATGTCGTCGCCGCTGAGCGTGCAGTGGCTCATGAGACGGGAAGAA	2700
Qy	2329	CAGATATTTGGAGCAAGTTCGTGAAATTCGTCCAGAGAAATTCGCAGAA--	2378
Db	2701	CAGATATTTGGAGCAAGTTCGTGAAATTCGTCCAGAGAAATTCGCAGAGAAATTCGCAGAAATTCGAGTTGTC	2760
Qy	2379	-----AAACAACGAC	2387
Db	2761	TTGAGTATTTAAAGATCTCTGGATTTTTTAAATTTTTTTTGTGAACTTTTCAGAACACAGAC	2820
Qy	2388	AATCGAGCACGCGCATATTTAGTCTATTCGAGACGGAGTTAGCGNATTCGGAGATGCTTA	2447
Db	2821	AATCGAGCACGCGCATATTTAGTCTATTCGAGACGGAGTTAGCGNATTCGGAGATGCTTA	2880
Qy	2448	CGTCTTAGTCATGATGAGCTTCGNATCTTTAAAGCGAAGTAAACAAATTCATGTCGGAA	2507
Db	2881	CGTCTTAGTCATGATGAGCTTCGNATCTTTAAAGCGAAGTAAACAAATTCATGTCGGAA	2940
Qy	2508	CGGGATGGAGAAGATCCAGAGCCGAAGTACACGTTCTATTCGATTCAGAAAAGACACAAT	2567
Db	2941	CGGGATGGAGAAGATCCAGAGCCGAAGTACACGTTCTATTCGATTCAGAAAAGACACAAT	3000
Qy	2568	ACACGNTGCTTCGAAGATGGAAAAAGATAAGCCAGTGTGTCAATAAAGATCTTACTCCT	2627
Db	3001	ACACGNTGCTTCGAAGATGGAAAAAGATAAAGCCAGTGTGTCAATAAAGATCTTACTCCT	3060
Qy	2628	GCTCAAAACAGATGTCGCTGTTGTCGTCTGTTTAAACAAATGGCAGGAGGATATGAAGAAGAAC	2687
Db	3061	GCTCAAAACAGATGTCGCTGTTGTCGTCTGTTTAAACAAATGGCAGGAGGATATGAAGAAGAAC	3120
Qy	2688	AAAGAACTGGAATTTGAAACCCATCATCCGGAACAACTCTGGATAACTTATCGTTTCG	2747
Db	3121	AAAGAACTGGAATTTGAAACCCATCATCCGGAACAACTCTGGATAACTTATCGTTTCG	3180

QY	2748	AAATACAAAATTCGAATTTTCTTGGCAATCTCATCATAGTGTCTCTTGGTGACATCTCGTCCA	2807
DB	3181	AAATACAAAATTCGAATTTTCTTGGCAATCTCATCATAGTGTCTCTTGGTGACATCTCGTCCA	3240
QY	2808	GGCATTACACTGTTATGTATGACGATAAAGGAATGAGCCAAAGATGAAGTCTAT----	2861
DB	3241	GGCATTACACTGTTATGTATGACGATAAAGGAATGAGCCAAAGATGAAGTCTATGTAAGC	3300
QY	2862	-----	2861
DB	3301	GTTTTGAATAGCAGTTAGCGATTTTAGGATTTTGTAAATCCGCATATAGTTATTATAAAAA	3360
QY	2862	-----AAATGACCTACGGACTTGCTTTCTCTCTGCTAGATGTCGAAAAACCCAT	2911
DB	3361	AATGTTTCAGAAAATGACCTACGGACTTGCTTTCTCTCTGCTAGATGTCGAAAAACCCAT	3420
QY	2912	CTCGTTCGCTGTTCCGGTTCATTATGCTCAATTTATCATGTGAAAAAGCAGAAAGAGCTTTA	2971
DB	3421	CTCGTTCGCTGTTCCGGTTCATTATGCTCAATTTATCATGTGAAAAAGCAGAAAGAGCTTTA	3480
QY	2972	TCGAACTTACAGGAACATTTACATCGGTGTACTATGCACAGCCACGGACTCGACACGAAT	3031
DB	3481	TCGAACTTACAGGAACATTTACATCGGTGTACTATGCACAGCCACGGACTCGACACGAAT	3540
QY	3032	GGAACATTTTCTCCAAACTAACCGTGAAGTACCCTGGAAATCTCGTTCGCATAACATTTTGC	3091
DB	3541	GGAACATTTTCTCCAAACTAACCGTGAAGTACCCTGGAAATCTCGTTCGCATAACATTTTGC	3600
QY	3092	AAAAGTGTGCGCGTTTCAATCAAAATTTTTTCAATCTGTAGATATTGTACTACTTTTTTTTT	3151
DB	3601	AAAAGTGTGCGCGTTTCAATCAAAATTTTTTCAATCTGTAGATATTGTACTACTTTTTTTTT	3660
QY	3152	AAAGCCCGGTTTCAAAAATTTCAATTCACATGACTAACCGTTTTTCATAAAATTACTTGAAATTT	3210
DB	3661	AAAGCCCGGTTTCAAAAATTTCAATTCACATGACTAACCGTTTTTCATAAAATTACTTGAAATTT	3719

## RESULT 6

```

US-11-144-985-1
; Sequence 1, Application US/11144985
; Publication No. US20060024798A1
; GENERAL INFORMATION:
; APPLICANT: Mello, Craig C.
; APPLICANT: Tabara, Hiroaki
; APPLICANT: Grishok, Alla
; APPLICANT: Fire, Andrew
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC
; TITLE OF INVENTION: RNA INTERFERENCE
; FILE REFERENCE: UMY-052CN
; CURRENT APPLICATION NUMBER: US/11/144,985
; CURRENT FILING DATE: 2005-06-03
; PRIOR APPLICATION NUMBER: US 09/689,992
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3719
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-11-144-985-1

```

Query Match	80.6%	Score 2501	DB 16	Length 3719
Best Local Similarity	86.3%	Pred. No. 0		
Matches 3210	Conservative	0	Mismatches 0	Indels 509
Gaps				10
1	CAGGCACAAAGTGATGAAACATGCTCGAAATTTTCCGAAATTCGAAAAAGGATTTTATC	60		
1	CAGGCACAAAGTGATGAAACATGCTCGAAATTTTCCGAAATTCGAAAAAGGATTTTATC	60		



QY 61 GTCATCTCTCGATCC----- 76  
Db 61 GTCATCTCTCGATCCGGTATGATCAATATTAGCAGCTATAAGATATAAAGTTTGATA 120  
QY 77 -----GGAGATGAATGGCTTGCAGAGCCCACTGGTAAATCGCAGCGCAAAATTC 125  
Db 121 TTAATATTATAGGAGATGAATGGCTTGCAGAGCCCACTGGTAAATCGCAGCGCAAAATTC 180  
QY 126 TATGAGAGAAAGTACTTCTTTTGGTAAATGGTTCAAGTTCTCCAGCAAAATTTACGAT 185  
Db 181 TATGAGAGAAAGTACTTCTTTTGGTAAATGGTTCAAGTTCTCCAGCAAAATTTACGAT 240  
QY 186 CGGGAATACTACGAGTATGAAGTGAATAACACAAAGGAAGTATTGAAATAGAAACACAGGA 245  
Db 241 CGGGAATACTACGAGTATGAAGTGAATAACACAAAGGAAGTATTGAAATAGAAACACAGGA 300  
QY 246 AAACCTTTCCAAAAAGACAGAAATTCCAAT----- 277  
Db 301 AAACCTTTCCAAAAAGACAGAAATTCCAATGTAAGTCTGTAAATTAGTCAAAACTA 360  
QY 278 -----TCCGATCGTGCAAAACTCTTCTGGCAACAATCTTGGCGCATGAGAAG 323  
Db 361 ATTATTATTTTCAGTCCCGATCGTGCAAAACTCTTCTGGCAACAATCTTGGCATGAGAAG 420  
QY 324 AAGCAGACAGATTTTATTTCTCGAAGACTATGTTTTGATGAAAGGACACTGTTTATAGT 383  
Db 421 AAGCAGACAGATTTTATTTCTCGAAGACTATGTTTTGATGAAAGGACACTGTTTATAGT 480  
QY 384 GTTTGTGCGACTGAACACTGTCAATCAAAATCTGTTTTCGGAGAAAGTAGTAAAAAAG 443  
Db 481 GTTTGTGCGACTGAACACTGTCAATCAAAATCTGTTTTCGGAGAAAGTAGTAAAAAAG 540  
QY 444 GATTCGGAGAAAGAAAGATGAAAGAGATTTGGAGAAAAAATCTTATACAAATGATACTT 503  
Db 541 GATTCGGAGAAAGAAAGATGAAAGAGATTTGGAGAAAAAATCTTATACAAATGATACTT 600  
QY 504 ACCATCGTAAATAATTTTCCCTGAACTTTAGTCGAGAAATCCGGAAGAACGAGAA 563  
Db 601 ACCATCGTAAATAATTTTCCCTGAACTTTAGTCGAGAAATCCGGAAGAACGAGAA 660  
QY 564 GCGAATCGGAGTTACAAATTCCTGA----- 588  
Db 661 GCGAATCGGAGTTACAAATTCCTGAGGTTTATGAAAAACACGCATTTATAATTTTAA 720  
QY 589 -----AGATGTTATGACCCAGAAAAGTTCGCTACCGGCCTTTTGTGAACGAGGAGA 639  
Db 721 TTAGCTTTCAGAAATGTTATGACCAGAAAAGTTCGCTACCGGCCTTTTGTGAACGAGGAGA 780  
QY 640 TTAAAGT----- 646  
Db 781 TTAAGTGTGAGTTGCAATAATAATAATCAATCACTCACTCAATTTATATATTTTAA 840  
QY 647 -ACAAATTCGCAAAAAATTTGTGTACGATAATAATCAATTTCTGCGAGTTCTCTGAATCGT 705  
Db 841 GACAATTCGCAAAAAATTTGTGTACGATAATAATCAATTTCTGCGAGTTCTCTGAATCGT 900  
QY 706 TTCAAGATCCAAACAGATTCGAACAAATCAATTAGAAGTAGCAACCAAGAAATCGAAGCATGTT 765  
Db 901 TTCAGCATCCAAACAGATTCGAACAAATCAATTAGAAGTAGCAACCAAGAAATCGAAGCATGTT 960  
QY 766 TTGGAATTTACATTGGAATCAAGAAATTTGTCAGTGTGAACTGTGCTCAATTTTGCAA 825  
Db 961 TTGGAATTTACATTGGAATCAAGAAATTTGTCAGTGTGAACTGTGCTCAATTTTGCAA 1020  
QY 826 -----TTGTGCAATAACT 838  
Db 1021 GTAAGTTTGAGAACTGCGATAAAAAATCATGTGATTTTGTGGAAGTTGTGATTAATCT 1080  
QY 839 ATTCTCAATGCAACGAAAAATGTTCTCTTCGGATTTATCTTCTCAATTTGTGACCCCA 898  
Db 1081 ATTCTCAATGCAACGAAAAATGTTCTCTTCGGATTTATCTTCTCAATTTGTGACCCCA 1140  
QY 899 GTCGTGTACGATGATGACAAAGATCTTAAACAAAACTGATCGCGGGAATAATGAC 958

Db 1141 GTCGTGTACGATGATGACAAAGATCTTAAACAAAACTGATCGCGGAATAATGAC 1200  
QY 959 AATCAGACAAGCCGCGCCAAAGAAATTCGACAAATATTTCGAAATTTTGAAGCTGAAATG 1018  
Db 1201 AATCAGACAAGCCGCGCCCAAGAAATTCGACAAATATTTCGAAATTTTGAAGCTGAAATG 1260  
QY 1019 CGCAGAAGTTTGGGATAACGAAAT----- 1042  
Db 1261 CGCAGAAGTTTGGGATAACGAAATGTTAGTTTAAATTTATTTCAAACTAATAATACAAAT 1320  
QY 1043 -----GTCCGAGATTGACAGAACGACATCTGACATTTCTAGATTTCTGCGAGGAAA 1092  
Db 1321 TGATTTTCAGGTCGAGATTGACAGAACGACATCTGACATTTCTAGATTTCTGCGAGGAAA 1380  
QY 1093 ACTCTCTCTGTTTATAAAGTCACTGGTAAATTCGACAGAGGAGAAATTCGCAAAAGATGACG 1152  
Db 1381 ACTCTCTCTGTTTATAAAGTCACTGGTAAATTCGACAGAGGAGAAATTCGCAAAAGATGACG 1440  
QY 1153 ATACTACATTTGTTCAAAATCTATGAGGAAAAAATAAAGTTTCATGTGATTTCCCCACCTAC 1212  
Db 1441 ATACTACATTTGTTCAAAATCTATGAGGAAAAAATAAAGTTTCATGTGATTTCCCCACCTAC 1500  
QY 1213 CACTAGTCAAGTTTAAAGTGGAGCAAAAGAAATACGCTGTACCAATGGACATCTTTGAAG 1272  
Db 1501 CACTAGTCAAGTTTAAAGTGGAGCAAAAGAAATACGCTGTACCAATGGACATCTTTGAAG 1560  
QY 1273 TTTCATGAGAAAGCCACAAAGATACAAAGATTCGAATTCGATCTGGTGATGCAAGACAAGTTTC 1332  
Db 1561 TTTCATGAGAAAGCCACAAAGATACAAAGATTCGAATTCGATCTGGTGATGCAAGACAAGTTTC 1620  
QY 1333 TAAAGCGAGCTACACGAAAACTTCACGACTACAAAGAAATAACCTTAAATAATGCTGAAAG 1392  
Db 1621 TAAAGCGAGCTACACGAAAACTTCACGACTACAAAGAAATAACCTTAAATAATGCTGAAAG 1680  
QY 1393 AATTGGAATTTCTCTCTGAGAGCTAAAATTTTGTGAAAGATTTGGAATTTGCTTCCCAAC 1452  
Db 1681 AATTGGAATTTCTCTCTGAGAGCTAAAATTTTGTGAAAGATTTGGAATTTGCTTCCCAAC 1740  
QY 1453 TTTCAGATGATCGAATGTCAGGAAAGGTTTGAAGAGGCAATGCTTGTGAATAGTGTAA 1512  
Db 1741 TTTCAGATGATCGAATGTCAGGAAAGGTTTGAAGAGGCAATGCTTGTGAATAGTGTAA 1800  
QY 1513 ATGAACAAATTTAAATGACACCAAGTGAATTCGTGGATTTCAAGAAATAACAAATGAAATGTGG 1572  
Db 1801 ATGAACAAATTTAAATGACACCAAGTGAATTCGTGGATTTCAAGAAATAACAAATGAAATGTGG 1860  
QY 1573 TTCCCGAAAAAGAACTTTTCTGCTGCTGTTTGTAGTCAACGAAACAGCGGGAATTCAT 1632  
Db 1861 TTCCCGAAAAAGAACTTTTCTGCTGCTGTTTGTAGTCAACGAAACAGCGGGAATTCAT 1920  
QY 1633 GCTTAGAAGAGAACGACGTTGT----- 1654  
Db 1921 GCTTAGAAGAGAACGACGTTGTGTAGTGTGTTTCTACGTAGATTATTCGAAATAATTTTC 1980  
QY 1655 ---TAAGTTCTACACCGAACTAATTTGGTGTGTCAGAGTTCCGTGGAAATACGAATGTGGTCC 1712  
Db 1981 AGTAAGTTCTACACCGAACTAATTTGGTGTGTCAGAGTTCCGTGGAAATACGAATGTGGTCC 2040  
QY 1713 AATGAACAGAGGAGCGCAATCTATTATGTCAGCGCGCAAGAAATGAAATATGCTCC--- 1769  
Db 2041 AATGAACAGAGGAGCGCAATCTATTATGTCAGCGCGCAAGAAATGAAATATGCTCCGTA 2100  
QY 1770 -----TTCTACAAAAATTTGTACAC 1788  
Db 2101 AGTTTCAGAAATTTGAAAGTTTAAATATCATATTATTCAGTTCTACAAAAATTTGTACAC 2160  
QY 1789 TAAATAACCGGAATTCGGTAGATTGGAATAGCCGCAACAGAGCGAAGAAATATGTTTGAAC 1848  
Db 2161 TAAATAACCGGAATTCGGTAGATTGGAATAGCCGCAACAGAGCGAAGAAATATGTTTGAAC 2220  
QY 1849 GTCCTCCGATTAAGAACAAAAAGTCTTAATGTTTCAATTCATTCATTTCCAAACGCAACTGA 1908

Db 2221 GTCTTCCGATAAAGAAACAAAAGTCTTAATGTTCAATTATCATTTCCAAACGACAACTGA 2280  
QY 1909 ATGCTTACCGTTTTGTGAAACATTAATTGCGATCAACCAATCGGTGTAGCTAATTCAGCATPA 1968  
Db 2281 ATGCTTACCGTTTTGTGAAACATTAATTGCGATCAACCAATCGGTGTAGCTAATTCAGCATPA 2340  
QY 1969 TTACTTCTCAACACAGTCACAAAGCTTTTGGCATCACTAAGGACGAGAAAGGATCAAAAC 2028  
Db 2341 TTACTTCTCAACACAGTCACAAAGCTTTTGGCATCACTAAGGACGAGAAAGGATCAAAAC 2400  
QY 2029 GAATTTTCTATCAAAATTGCAATTTGAAATCAACGCGAAATTTAGGAGGTATTAACAGGAGC 2088  
Db 2401 GAATTTTCTATCAAAATTGCAATTTGAAATCAACGCGAAATTTAGGAGGTATTAACAGGAGC 2460  
QY 2089 TTGACTGGTTCAGAAATTCAGAAATATCAACGAGAAAGAAAGAACGCGAAACAAATGC 2148  
Db 2461 TTGACTGGTTCAGAAATTCAGAAATATCAACGAGAAAGAAAGAACGCGAAACAAATGC 2520  
QY 2149 CATTAACTATGTATGTGGAATTTGATTAACCTCACTCAACCTCCTACAGTGGAAATTTGATT 2208  
Db 2521 CATTAACTATGTATGTGGAATTTGATTAACCTCACTCAACCTCCTACAGTGGAAATTTGATT 2580  
QY 2209 ATTCTATAGCGGTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGA 2268  
Db 2581 ATTCTATAGCGGTGTAGTAGCGAGTATCAATCCAGGTGGAACTATCTATCGAAATATGA 2640  
QY 2269 TTGTGACTCAAGAGAAATGTGTCGCGGTGAGGTGCGATGCGTCAATGACGCGGAAAGAA 2328  
Db 2641 TTGTGACTCAAGAGAAATGTGTCGCGGTGAGGTGCGATGCGTCAATGACGCGGAAAGAA 2700  
QY 2329 CAGATATTTTGAAGCAAAAGTTTGTGAAATTTGCTCAGAGAAATTCGAGAA----- 2378  
Db 2701 CAGATATTTTGAAGCAAAAGTTTGTGAAATTTGCTCAGAGAAATTCGAGAAATTCGAGAA 2760  
QY 2379 -----AACACGAC 2387  
Db 2761 TTGAGTATTTAAAGATCTCGGATTTTAAATTTTGTGTAACCTTTTCAGAAACACGAC 2820  
QY 2388 AATCGAGACACGCGCATATTTAGTCTATTCGAGACGGAGTTAGCGATTCGGAGATGCTPA 2447  
Db 2821 AATCGAGACACGCGCATATTTAGTCTATTCGAGACGGAGTTAGCGATTCGGAGATGCTPA 2880  
QY 2448 CGTGTGTAGTCATGATGAGCTTCGATCTTTTAAAGCGAAGTAAACAAATTCATGTCGAA 2507  
Db 2881 CGTGTGTAGTCATGATGAGCTTCGATCTTTTAAAGCGAAGTAAACAAATTCATGTCGAA 2940  
QY 2508 CGGGATGGAGAGATCCAGAGCGGAAGTACAGTTCATTTGTGATTCAGAAAGACACAAAT 2567  
Db 2941 CGGGATGGAGAGATCCAGAGCGGAAGTACAGTTCATTTGTGATTCAGAAAGACACAAAT 3000  
QY 2568 ACACGATTCCTCGAAGATGGAAAGATAGACCGAGTGGTCAATAAAGATCTTACTCT 2627  
Db 3001 ACACGATTCCTCGAAGATGGAAAGATAGACCGAGTGGTCAATAAAGATCTTACTCT 3060  
QY 2628 GCTGAAACAGATGTCGCTGTGCTGTTTAAACATGGAGGAGGATAGAAAGAAC 2687  
Db 3061 GCTGAAACAGATGTCGCTGTGCTGTTTAAACATGGAGGAGGATAGAAAGAAC 3120  
QY 2688 AAAGAACTGGAAATTTGTGAACCCATCATCCGAAACAACTGTGGATAAACTTATCGTTTTCG 2747  
Db 3121 AAAGAACTGGAAATTTGTGAACCCATCATCCGAAACAACTGTGGATAAACTTATCGTTTTCG 3180  
QY 2748 AAATACAAATTCGATTTTTTTCTTGGCATCTCATGCTGCTTGGTACATCTCGTCCA 2807  
Db 3181 AAATACAAATTCGATTTTTTTCTTGGCATCTCATGCTGCTTGGTACATCTCGTCCA 3240  
QY 2808 GGACATTTACACTGTTATGATGACGATAAGGAAATGAGCAAGATCAAGTCTAT----- 2861  
Db 3241 GGACATTTACACTGTTATGATGACGATAAGGAAATGAGCAAGATCAAGTCTATGTAAGC 3300  
QY 2862 ----- 2861  
Db 3301 GTTTTGAAATAGCAGTTAGCGATTTTGGATTTTGTAAATCCGCATATAGTTATTATAAAAA 3360

RESULT 7

US-10-174-363-3  
; Sequence 3, Application US/10174363  
; Publication No. US20030077623A1  
; GENERAL INFORMATION:  
; APPLICANT: Butler, Karlene H.  
; APPLICANT: Cahoon, Rebecca E  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Rafalski, Antoni J.  
; APPLICANT: Sakai, Hajime  
; TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptiona  
; TITLE OF INVENTION: Gene Silencing  
; FILE REFERENCE: BB1454 US NA  
; CURRENT APPLICATION NUMBER: US/10/174,363  
; PRIOR FILING DATE: 2002-06-17  
; CURRENT APPLICATION NUMBER: 60/298,973  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 3  
; LENGTH: 2827  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-174-363-3

Query Match 2.0%; Score 63.8; DB 6; Length 2827;

Best Local Similarity 52.0%; Pred. No. 0.00026;

Matches 143; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 2674 ATATCAAGAAGACGAAGAACTGGAAATTTGCAACCCATCATCCGAAACAACTGTGCGATA 2733  
Db 2028 ACAGGACAGAGAGATGACAGATCGAGTGGGAATATATTGCTCGGACTGTGTTGATA 2087  
QY 2734 AACTTATCGTTTTCGAAATACAAATTCGATTTTTTTCTTGGCATCTCATATGCTGCTTG 2793  
Db 2088 CCAAAATCTGCCATCCAACAGAAATTTGATTTTTTATCTCTGCAGCCATGCTGGCATCCAG 2147  
QY 2794 GTACATCTCGTCAGACATTTACACTGTATGATGACGATAAGGAATAGCCAGATG 2853  
Db 2148 GTACTAGTCGGCAGCTCATTTATGCTCTGCTGGGATGAAACAACTTCACACCTGATG 2207  
QY 2854 AAGTCTATAAAATGACCTACGAGCTTGTCTTCTCTGCTAGATGTCGAAACCCATCT 2913  
Db 2208 GAATTCAGTCTCTGACAAACACCTTTGTTATACATATGCCAGGTGTACACGCTCAGTAT 2267  
QY 2914 CGTTCCTGTTCCGGTTCAATTATGCTCAATTTATCA 2948



Db 2268 CAGTTGTTCTCCAGCATATATTATGCACATTTAGCA 2302

RESULT 8

US-11-093-888-3

Sequence 3, Application US/11093888

Publication No. US20050204427A1

GENERAL INFORMATION:

APPLICANT: Butler, Karlene H.

APPLICANT: Cahoon, Rebecca E

APPLICANT: Harvelli, Leslie T.

APPLICANT: Rafalski, Antoni J.

APPLICANT: Sakai, Hajime

TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptional

FILE REFERENCE: B01454 US NA

CURRENT APPLICATION NUMBER: US/11/093,888

CURRENT FILING DATE: 2005-03-30

PRIOR APPLICATION NUMBER: US/10/174,363

PRIOR FILING DATE: 2002-06-17

PRIOR FILING DATE: 2001-06-18

NUMBER OF SEQ ID NOS: 56

SOFTWARE: Microsoft Office 97

SEQ ID NO 3

LENGTH: 2827

TYPE: DNA

ORGANISM: Glycine max

US-11-093-888-3

Query Match 2.0%; Score 63.8; DB 13; Length 2827;

Best Local Similarity 52.0%; Pred. No. 0.00026;

Matches 143; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 2674 ATATGAAGAAGCAAGCAAACTGGAATTTGTGAACCCATCATCGGAAACAACTGTGGATA 2733

Db 2028 ACAGGACAGACAGCAGTACAGATCGGAGTGGGAATATATTGCTGGGACATGTTGGATA 2087

QY 2734 AACTATCGTTTGAATACAAATTCGATTTTCTGTTGGCATCTCATCATGTTGCTCTTG 2793

Db 2088 CCAAAATTCGCCATCCCAACAGAAATTTGATTTTATCTCTGCAGCCATCTGGCATCCAGG 2147

QY 2794 GTACATCTCTCCAGGACATTTACATCTGTATGTATGATGAGTAAGGAATGAGCCAGATG 2853

Db 2148 GTACTAGTCGGCAGCTCATTCATGCTCTGTGGGATGAACAACTTCACACCTGATG 2207

QY 2854 AAGCTTAAATGACCTACGAGCTTGTCTTCTCTGTAGATGCGAAACCCATCT 2913

Db 2208 GAATTCAGTCTCTGACAAACAACTTTGTTATACATATGCCAGTGTACAGCTCAGTAT 2267

QY 2914 CGTTGCTGTTCGGTTTCATTTATGCTCATTTATCA 2948

Db 2268 CAGTTGTTCTCCAGCATATATTATGCACATTTAGCA 2302

RESULT 9

US-09-918-995-10623

Sequence 10623, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 10623

LENGTH: 490

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)...(490)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-10623

Query Match 1.9%; Score 62.4; DB 3; Length 490;

Best Local Similarity 51.8%; Pred. No. 0.00021;

Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 2698 GAATTGTGAACCCCATCATCGGAACTGTGGATAAACTTATCGTTTCGAAATACAAAT 2757

Db 215 GAATGTCATATCCAGCTGGAACAAGTTGTATACAGATTACACACCATATGAGT 274

QY 2758 TCGATTTTCTCTGTCATCATCGTGTCTTGGTGTATCTCTGTCAGGACATTTACA 2817

Db 275 TCGATTTTCTCTGTCATCATCGTGTCTTGGTGTATCTCTGTCAGGACATTTACA 334

QY 2818 CTGTTATGATGAGTAAAGGAATGAGCAAGATGAAGTCTATAAATGACCTACGAGC 2877

Db 335 ATGTTTATGGGATGATACTGCTTTACTGCAGATGAACCTTCAGCTGCTAACTTACCAGC 394

QY 2878 TTGCTTTTCTCTGTCATCATCGTGTCTTGGTGTATCTCTGTCAGGACATTTACA 2937

Db 395 TCTGCCACACTTACGATGCTGTACAGATCTGTTTCTATACCTGCACCCAGCGTATTATG 454

QY 2938 CTCATTTTATCATGTGAAAAAGCGAAAGAGCTT 2969

Db 455 CTCACCTGTAGCATTAGAGCCAGATATCAT 486

RESULT 10

US-10-779-543-3372

Sequence 3372, Application US/10779543

Publication No. US20050227917A1

GENERAL INFORMATION:

APPLICANT: Williams et al

TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED

TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II

FILE REFERENCE: 2300-21302

CURRENT APPLICATION NUMBER: US/10/779,543

CURRENT FILING DATE: 2004-02-12

PRIOR APPLICATION NUMBER: 10/076,555

PRIOR FILING DATE: 2002-02-15

PRIOR APPLICATION NUMBER: 09/217,471

PRIOR FILING DATE: 1998-12-21

PRIOR APPLICATION NUMBER: 60/068,755

PRIOR FILING DATE: 1997-12-23

PRIOR APPLICATION NUMBER: 60/080,664

PRIOR FILING DATE: 1998-04-03

PRIOR APPLICATION NUMBER: 60/105,234

PRIOR FILING DATE: 1998-10-21

PRIOR APPLICATION NUMBER: 09/297,648

PRIOR FILING DATE: 2000-04-10

PRIOR APPLICATION NUMBER: PCT/US99/01619

PRIOR FILING DATE: 1999-01-28

PRIOR APPLICATION NUMBER: 60/072,910

PRIOR FILING DATE: 1998-01-28

PRIOR APPLICATION NUMBER: 60/075,954

PRIOR FILING DATE: 1998-02-24

PRIOR APPLICATION NUMBER: 60/080,114

PRIOR FILING DATE: 1998-03-31

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 23767

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3372

LENGTH: 734

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 5, 6, 7, 8, 9, 10, 26, 639





Db	971	GTCTGTGGACGAGAACAAATTTCAC	GTGCTGATGAGATCCATCTCTGACCACTTG	1030
Qy	2880	GCTTTTCTCTGCTAGATGTCGAAAA	CCCATCTCGTTCCTGTTCCGGTTCATTATGCT	2939
Db	1031	TGCTACACCTATGCAAGATGTACAC	CGATCAGTTTCTGTAGTGCTCTGCGTACTATGCT	1090
Qy	2940	CATTTATCATGTGAAAAAGC	2959	
Db	1091	CATTGGCAGCTTACAGAGC	1110	

Search completed: July 5, 2006, 23:14:00  
Job time : 3768 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2006, 22:10:12 ; Search time 385 Seconds  
(without alignments)  
10897.477 Million cell updates/sec

Title: US-10-645-746-2  
Perfect score: 3227  
Sequence: 1 cagccacaagtgtgaac.....tttaaaaaaaaaaaaaa 3227

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 844069 seqs, 650066433 residues

Total number of hits satisfying chosen parameters: 1688138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:  
1: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
2: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
3: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
4: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
5: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
6: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
7: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
8: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
9: /EMC Cellerra\_SIDS3/ptodata/1/pubpna/US60\_NEW\_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62.4	1.9	3320	8	US-11-266-748A-31508
2	62.4	1.9	3320	8	US-11-266-748A-31508
3	61	1.9	1196	6	US-10-449-902-3167
4	61	1.9	1196	6	US-10-449-902-3167
5	61	1.9	3178	6	US-10-449-902-3806
6	56.4	1.7	4758	8	US-11-266-748A-30266
7	54.6	1.7	1211	6	US-10-449-902-26400
8	54	1.7	1094	7	US-11-218-305-21995
9	53.6	1.7	892	7	US-11-218-305-19386
10	52.8	1.6	3350	7	US-11-218-305-19388
11	52	1.6	1710	7	US-11-218-305-19385
12	51	1.6	3474	6	US-10-953-349-4063
13	50.2	1.6	3422	6	US-10-449-902-27837
14	46.6	1.4	964	7	US-11-218-305-22488
15	46.6	1.4	965	7	US-11-218-305-22488
16	46.6	1.4	2001	7	US-11-218-305-22483
17	46.2	1.4	1816	6	US-10-953-349-19397
18	45.8	1.4	1469	7	US-11-218-305-22082
19	45.6	1.4	3394	8	US-11-266-748A-26618
20	45.6	1.4	3394	8	US-11-266-748A-29538
21	45.2	1.4	2016	7	US-11-218-305-22484
22	45.2	1.4	3483	6	US-10-449-902-26806
23	45.2	1.4	3746	7	US-11-218-305-22486
24	45	1.4	997	8	US-11-216-545-6851

25	44.6	1.4	1793	8	US-11-266-748A-25558	Sequence 25558, A
c 26	44.4	1.4	3397	6	US-10-517-441-378	Sequence 378, App
c 27	44.4	1.4	3397	6	US-10-517-441-378	Sequence 652, App
28	44.2	1.4	1000	8	US-11-266-748A-394325	Sequence 394325, A
c 29	44.2	1.4	1000	8	US-11-266-748A-465371	Sequence 465371, A
30	44.2	1.4	7478	8	US-11-266-748A-56189	Sequence 56189, A
31	43.6	1.4	903	7	US-11-218-305-19468	Sequence 19468, A
32	43.6	1.4	2095	7	US-11-218-305-1556	Sequence 1556, App
33	43.2	1.3	1555	8	US-11-216-545-7032	Sequence 7032, App
34	43	1.3	1934	8	US-11-216-545-6852	Sequence 6852, App
35	42.8	1.3	2773	7	US-11-218-305-19467	Sequence 19467, A
36	42.4	1.3	2503	6	US-10-449-902-19320	Sequence 19320, A
37	41.8	1.3	1000	8	US-11-266-748A-290582	Sequence 290582, A
c 38	41.8	1.3	1000	8	US-11-266-748A-342011	Sequence 342011, A
39	41.8	1.3	1000	8	US-11-266-748A-401890	Sequence 401890, A
c 40	41.8	1.3	1000	8	US-11-266-748A-472936	Sequence 472936, A
41	41.8	1.3	294540	8	US-11-266-748A-23953	Sequence 23953, A
42	40.8	1.3	951	8	US-11-266-748A-48160	Sequence 48160, A
c 43	40.8	1.3	1298	8	US-11-216-545-2984	Sequence 2984, App
c 44	40.6	1.3	713	8	US-11-266-748A-211284	Sequence 211284, A
c 45	40.6	1.3	713	8	US-11-266-748A-234991	Sequence 234991, A

## ALIGNMENTS

RESULT 1  
US-11-266-748A-31508  
; Sequence 31508, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and Methods of Using the Same  
; FILE REFERENCE: 5815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; PRIOR FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 483996  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 31508  
; LENGTH: 3320  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-11-266-748A-31508

Query Match 1.9%; Score 62.4; DB 8; Length 3320;  
Best Local Similarity 51.8%; Pred. No. 3.4e-05;  
Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
QY 2698 GAATTGTGAACCCATCATCGGAACAACTGTGGATAAACTTATCGTTTGAATAACAAT 2757  
Db 2178 GAAGTGGCAATATCCAGCTGGGAACAACAGTTGATACAGACATTACACACCATATGAGT 2237  
QY 2758 TCGATTTCCTTGGCATCTCATCGTGTCTCTGTGTACATCTCGTCCAGGACATTACA 2817

Db 2238 TCGATTTTACCTCTGTAGCCATGCTGGAATACAGGGTACCAGTCTGCTTCACACTATC 2297  
Qy 2818 CTGTTATGATGACGATAAGGAATGAGCCAGATGAAGTCTATAAATGACCTACGGAC 2877  
Db 2298 ATGTTTTATGGGATGATACTGCTTTACTGACAGTGAACCTTCAGCTGCTAACTTACCAGC 2357  
Qy 2878 TTGCTTTTCTCTGCTAGATGCGAAACCCCATCTCGTTGCTGTTCCGGTTCATTATG 2937  
Db 2358 TCTGCCACACTTACGTACGCTGTACACGATCTGTTTCTATACCTGCACCGGTATTATG 2417  
Qy 2938 CTCATTATCATGTGAAAAGCGAAAGAGCTT 2969  
Db 2418 CTCACCTGGTAGCTTTAGAGCCAGATATCAT 2449

## RESULT 2

US-11-266-748A-56999  
; Sequence 56999, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkin, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcriptome Microarray Technology and  
; METHOD OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; CURRENT FILING DATE: 2005-11-03  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 483996  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 56999  
; LENGTH: 3320  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-11-266-748A-56999

Query Match 1.9%; Score 62.4; DB 8; Length 3320;  
Best Local Similarity 51.8%; Pred. No. 3.4e-05;  
Matches 141; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
Qy 2698 GAATGTGAACCATCATCCGAGACAACTGTGATAAATCTTATCGTTTCAAAATCAAT 2757  
Db 2178 GAAGTGGCAATATCCAGCTGGAAACAACTGTTGATACAGACATTACACACCCATATGAGT 2237  
Qy 2758 TCGATTTTCTTGGCATCTCATATGTTGTTCTTGGTATCATCTCGTCCAGGACATTACA 2817  
Db 2238 TCGATTTTACCTCTGTAGCCATGCTGGAATACAGGTTACAGTCTGCTTTCACACTATC 2297  
Qy 2818 CTGTTATGATGACGATAAGGAATGAGCCAGATGAAGTCTATAAATGACCTACGGAC 2877  
Db 2298 ATGTTTTATGGGATGATACTGCTTTACTGACAGTGAACCTTCAGCTGCTAACTTACCAGC 2357  
Qy 2878 TTGCTTTTCTCTGCTAGATGTCGAAACCCCATCTCGTTGCTGTTCCGGTTCATTATG 2937  
Db 2358 TCTGCCACACTTACGTACGCTGTACACGATCTGTTTCTATACCTGCACCGGTATTATG 2417  
Qy 2938 CTCATTATCATGTGAAAAGCGAAAGAGCTT 2969

Db 2418 CTCACCTGGTAGCATTTAGAGCCAGATATCAT 2449

## RESULT 3

US-10-449-902-3167  
; Sequence 3167, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A020511-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3167  
; LENGTH: 1196  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: AK060907  
; DATABASE ENTRY DATE: 2001-12-06  
US-10-449-902-3167

Query Match 1.9%; Score 61; DB 6; Length 1196;

Best Local Similarity 53.1%; Pred. No. 4.7e-05;  
Matches 130; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
Qy 2716 CCGGAACCAACTGTGGATAAACTTATCGTTTCGAAATACAAATTCGATTTTCTTGGCAT 2775  
Db 517 CAGGTACTGTTGTGGACAAACAAAGTGTGCCATCCAGGAATATGACTTCTACATGTG 576  
Qy 2776 CTCATCATGGTGTCTTGGTACATCTCGTCCAGGACATTPACACTGTTATGTATGACGATA 2835  
Db 577 CTCATGCTGGAATGATTCGACGAGCCAAACATTTATCATGTCTTGCACGATGAGA 636  
Qy 2836 RAGGAATGAGCCAGATGAAGTCTATAAAATACACCTAGCGACTTCTTCTCTCTGCTA 2895  
Db 637 TAGGTTTCTCCCTCATGATCTGCAGGAGCTAGTGCACTCTCTTATGTATCAGA 696  
Qy 2896 GATGTCGAAAACCCCATCTCGTTGCTGCTTCCGTTTCATTATGCTCATTTATCATGTGAAA 2955  
Db 697 GNAGCACACAGCCATATCATGTTGTTGCTCCGATTTGCTATGCCATCTTGTCTGCTCTC 756  
Qy 2956 AAGCG 2960  
Db 757 AGGTG 761

## RESULT 4

US-10-449-902-3806  
; Sequence 3806, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A020511-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870

```

; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3806
; LENGTH: 1196
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: A
; DATABASE ENTRY DATE: 2002-08-
US-10-449-902-3806

```

**Query Match** 1.98; Score 61; DB 6; Length 1196;

Qy	2716	CCGGAACAACTGTGGATAAACTTTATCGTTTCGAAATACAAATTCGATTTTTTCTTGGCAT	2775
Db	517	CAGGTACTGTGTGGACAAACAAGTGTGCCATCCACAGGAATATATGACTTCTACATGTGTG	576
Qy	2776	CTCATCATGTGTCTCTTGGTACATCTCGTCCAGGACACATTACACTGTTATGTATGACACATA	2835
Db	577	CTCATGCTGGAATGATTGGAAACGACGAGGCCACACATTATCATGTTCTTGCAACGATGAGA	636
Qy	2836	AAGGAATGAGCCCAAGATGAAGTCTATAAATGACCTACGGACTGTCTTTCTCTCTGCTA	2895
Db	637	TAGGTTTCTCCCTGTGATGATCTGCAGGAGCTAGTGCACCTCACTCTCTTATGTGTATCAGA	696
Qy	2896	GATGTGGAACCAACCATCTCGTTGCTGTTCCGGTTCAATTATGCTCAATTTATCATGTGAAA	2955
Db	697	GAGACACACAGCCATATCAGTTGTCTCCGATTTCGTATGCCCATCTTGTGCTGCTGTC	756
Qy	2956	AAGCG	2960
Db	757	AGGTG	761

RESULT 5  
 US-10-449-902-14723  
 ; Sequence 14723, Application US/10449902  
 ; Publication No. US20060123505A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: National Institute of Agrobiological Sciences.  
 ; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
 ; APPLICANT: The Institute of Physical and Chemical Research.  
 ; APPLICANT: Foundation for Advancement of International Science.  
 ; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
 ; FILE REFERENCE: MOA-A0205V1-US  
 ; CURRENT APPLICATION NUMBER: US/10/449,902  
 ; CURRENT FILING DATE: 2003-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2002-203269  
 ; PRIOR FILING DATE: 2002-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2002-363870  
 ; PRIOR FILING DATE: 2002-12-11  
 ; NUMBER OF SEQ ID NOS: 56791  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 14723  
 ; LENGTH: 3178  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; PUBLICATION INFORMATION:  
 ; DATABASE ACCESSION NUMBER: AK065274  
 ; DATABASE ENTRY DATE: 2001-12-06  
 US-10-449-902-14723

	Query Match	1.9%	Score 61;	DB 6;	Length 3178;
	Best Local Similarity	53.1%;	Pred. No. 7.3e-05;		
	Matches 130;	Conservative 0;	Mismatches 115;	Indels 0;	Gaps 0;
QY	2716	CCGGAAACA	ACTGGGATAA	CTTATCGTTTCGAAATA	CAAAATTCGATTTTTCTTGCGAT 2775
DB	2476	CAGGTACTGTTGGG	CAAAACAAGTGTGCAT	CCGAAGAAATATGATCTTACATGTGTG	2535

Qy	2776	CTCATCATGGTGTCTCTTGGTACATCTCGTCCAGGACATTACACTGTTATGTATGACGATA	2835
Db	2536	CTCATGTGGAATGATTGGAACGACGAGGCCAACACATTTATCATGTTCTGCACGATGAGA	2595
Qy	2836	AAGGAATGAGCCAAAGATGAAGTCTATAAAATGACCTTAGCGACTGCTTTTCTCTCTGCTA	2895
Db	2596	TAGGTTTCTCCCTCGATGATCTGCGAGAGCTAGTGCACTCACTCTCTTAATGTGTATCAGA	2655
Qy	2896	GATCTGAAAACCCATCTCGTTGCTCTTCGGTTCATTATGCTCAATTTATCATGTGAAA	2955
Db	2656	GAAGCACACAGCCATATCAAGTTGTGCTCCGATTTGCTAATGCCCATCTTGTGCTGCTC	2715
Qy	2956	AAGCG	2960
Db	2716	AGGTG	2720

```

RESULT 6
US-11-266-748A-30266
; Sequence 30266, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcriptome Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266, 748A
; CURRENT FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662, 276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700, 293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 483996
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 30266
; LENGTH: 4758
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-30266

```

Query Match	1.7%	Score 56.4;	DB 8;	Length 4758;
Best Local Similarity	52.1%	Pred. No. 0.0042;		
Matches 126; Conservative	0;	Mismatches 116;	Indels 0;	Gaps 0;
Qy	2699	AATTTGTGAACCCCATCATCGGAAACAACTGTGGATAAATTTATCGTTTCGAAATACAAAT	2758	
Db	2374	AAGTGGCAATGTACAGCAGGCACTACAGTGGNTAGTACCTACACATCCATCTGAGTT	2433	
Qy	2759	CGATTTTTCTTTGGCANCTCATCATGGTGCCTTGGTATACATCTGTTCCAGGACATTACAC	2818	
Db	2434	TGACTTTTAACTCTGTGTAGTCATGTCAGGAAATTCAGGGAACGACGCCGTCCCTCACAATTACCA	2493	
Qy	2819	TGTTATGTATGACGATAAAGGAATGACCAAGATGAAGTCTATAAAATGCACTTACGGACT	2878	
Db	2494	GGTCTTGTGGATGACACTGCTTCACTCGAATGAATCCAGCTACTGACTTACAGCT	2553	
Qy	2879	TGCTTTTCTCTCTGTAGATGTGAAACCCCATCTCGTTTGCCTGTTTCGGTTTCATTATGC	2938	
Db	2554	GTGTACACCTATGTGAGGTGCACTCGCTCAGTCTTATTCAGGCCCTGCATATTATGC	2613	



Qy	2939	TC	2940
Db	2614	CC	2615

```

RESULT 7
US-10-449-902-26400
; Sequence 26400, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiol
; APPLICANT: Bio-oriented Technology Resea
; APPLICANT: The Institute of Physical and
; APPLICANT: Foundation for Advancement of
; TITLE OF INVENTION: FULL-LENGTH PLANT CDN
; FILE REFERENCE: MOA-A0205V1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 26400
; LENGTH: 1211
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK101842
; DATABASE ENTRY DATE: 2002-08-28
US-10-449-902-26400

```

Query Match	1.7%	Score 54.6	DB 6	Length 1211
Best Local Similarity	49.1%	Pred. No. 0.0018		
Matches 144	Conservative 0	Mismatches 149	Indels 0	Gaps 0
Qy	2656	TTAAACAATGGGAGGAGATGATGAAGAAAGCAAGAAACATGGAAATTTGTGAACCCATCAT	2715	
Db	493	TTGCTCAAAAGAAATCACCAACACAAAACCTTTTCCAAGAAATACTCCAGACAACGTTCCCTC	552	
Qy	2716	CCGGAAACACTGTGGATAAATTATCTTTTGGAAATACAAATTCGATTTTCTTGGCAT	2775	
Db	553	CAGGGACTGTTGTAGATCTCGTATTTGTCATCCAGACAGTAGATTTCTACATGATG	612	
Qy	2776	CTCATCATGGTGTCTTGGTACATCTCGTCCAGGACATTACACTGTTATGTATGACGATA	2835	
Db	613	CTCATGTGGACCTATAGTAGTACCTCAAGGCCACCCATTACCATTCTTGCTCGATGAGA	672	
Qy	2836	AAGGAATGACCCAGATGAAGTCTATAAAATGACCTACGGACTTGCTTTCTCTCTGCTA	2895	
Db	673	TTGGTTTCTTGGCCGATGACGTGAGAAACTAGTTTTGTGCGTTCTATGTGTATCAGA	732	
Qy	2896	GATGTGAAAAACCAATCTCGTTGCGCTGTTCCGGTTCAATTATGCTCATTTATCA	2948	
Db	733	GGAGCACCACTGCAATATCTGTTGGCAACCTATCTGTATGCTCATCTCGCA	785	

```

RESULT 8
US-11-218-305-21995/c
; Sequence 21995, Application US/11218305
; Publication No. US20060141495A1
; GENERAL INFORMATION:
; APPLICANT: MONSANTO TECHNOLOGY, LLC
; APPLICANT: Paul L.
; APPLICANT: McLaaird, Paul L.
; APPLICANT: Tao, Nengbing
; APPLICANT: Wu, Kunsheng
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping
; TITLE OF INVENTION: Corn.
; FILE REFERENCE: 38-21 (53660)B
; CURRENT APPLICATION NUMBER: US/11/218,305
; CURRENT FILING DATE: 2005-09-01

```

; PRIOR APPLICATION NUMBER:	US	60/606,880
; PRIOR FILING DATE:	2004-09-01	
; NUMBER OF SEQ ID NOS:	25043	
; SOFTWARE:	PatentIn version 3.2	
; SEQ ID NO	21995	
; LENGTH:	1094	
; TYPE:	DNA	
; ORGANISM:	Zea mays	
US-11-218-305-21995		
Query Match	1.7%; Score 54; DB 7; Length 1094;	
Best Local Similarity	51.2%; Pred.No. 0.0024;	
Matches 126; Conservative	0; Mismatches 120; Indels 0; Gaps 0;	
Qy	2696 TGAATTGTGAAACCCATCATCTCGGAACAACACTGTGGATAAACCCTTATCGTTTCGAAATACAA	2755
Db	764 TCGAAGTGGAACCAATTTTGCACGAACTGTGTGTGACACGAAGAATATGCCATCCAAGTGA	705
Qy	2756 ATTTCGATTTTTTCTTGCGCATCTCATCATGTGTGCTTTGGTACATCTCGTCCAGGACATTA	2815
Db	704 ATTCGACTTCTACTTTGTAGCACCCTCTGGCATACAGSGNACGAGCGGCCACACACTA	645
Qy	2816 CACTGTTATGTTATGACGATAAAGGAATGAGCCAAAGATGAAGTCTTATAAATGACCTTAGCGG	2875
Db	644 CCATGTTCTCTTTTGACGAAAAACAATTTCACTGCCGATGCACCTGCAAAACGTTGACCTTACAA	585
Qy	2876 ACTTGCTTTTTCTCTCTGCTAGATGTGCGAAAACCCATCTCGTTGCCCTGTTCCGGTTCATTA	2935
Db	584 ACTGTGCTACACGATATGCCGGATGCACCTCAGTCTCTATAGTTAGTTCCCCCGCGGTACTA	525
Qy	2936 TGCTCA	2941
Db	524 TGCACA	519

```

RESULT 9
US-11-218-305-19386
; Sequence 19386, Application US/11218305
; Publication No. US2006014195A1
; GENERAL INFORMATION:
; APPLICANT: MONSANTO TECHNOLOGY, LLC
; APPLICANT: McLaird, Paul L.
; APPLICANT: Tao, Nengbing
; APPLICANT: Wu, Kunsheng
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping
; TITLE OF INVENTION: Corn.
; FILE REFERENCE: 38-21 (536660) B
; CURRENT APPLICATION NUMBER: US/11/218,305
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US 60/606,880
; PRIOR FILING DATE: 2004-09-01
; NUMBER OF SEQ ID NOS: 25043
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19386
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Zea mays
US-11-218-305-19386

```

Db 358 AGATAGGCTTCAATCCTGATGACCTGCAGGAGCTGGTGCACTCGCTCTCTTATGTGTACC 417  
Qy 2893 CTGATGTCGAAACCCATCTGTTGCTGCTGTTCCGGTTCATTATGCTCATTTATCA 2948  
Db 418 AAAGGAGCACAAAGCCATATCATGTTGCTCCCATCTGCTAGCAGCACACCTTGCA 473

## RESULT 10

US-11-218-305-19388  
; Sequence 19388, Application US/11218305  
; Publication No. US2006014195A1  
; GENERAL INFORMATION:  
; APPLICANT: MONSANTO TECHNOLOGY, LLC  
; APPLICANT: McLaird, Paul L.  
; APPLICANT: tao, Nengbing  
; APPLICANT: Wu, Kunsheng  
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
; FILE REFERENCE: 38-21 (53660)B  
; CURRENT APPLICATION NUMBER: US/11/218,305  
; CURRENT FILING DATE: 2005-09-01  
; PRIOR APPLICATION NUMBER: US 60/606,880  
; NUMBER OF SEQ ID NOS: 25043  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 19388  
; LENGTH: 3350  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (738)..(742)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (908)..(930)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3235)..(3236)  
; OTHER INFORMATION: n is a, c, g, or t  
US-11-218-305-19388

Query Match 1.6%; Score 52.8; DB 7; Length 3350;  
Best Local Similarity 51.2%; Pred. No. 0.0077;  
Matches 123; Conservative 0; Mismatches 117; Indels 0; Gaps 0;  
Qy 2713 CATCCGGAACAACCTGTGGTAAACTTATCGTTTCGAAATACAAATTCGATTTTCTTGG 2772  
Db 2525 CACCAGAACTGGTGGACACAAAGCTGCGCATCCAAGAACTTCGATTTCTACATGT 2584  
Qy 2773 CATCTCATCATGTCCTTGGTACATCTCGTCCAGGACATTACACTGTTATGTATGACG 2832  
Db 2585 GTGGCATGCTGGAATGCTGGGACTACGAGGCACTCACTACCACATCCTGCATGATG 2644  
Qy 2833 ATAAAGGATGACCCAGATGAAGTCTATAAAATGACCTACGAGACTGCTTTCTCTCG 2892  
Db 2645 AGATAGGCTTCACTCCTCGATGATCTGAGGAGCTGGTGCATTCGCTCTCTTATGTACC 2704  
Qy 2893 CTAGATGTCGAAACCCATCTCGTTGCTTCCGGTTCATTATGCTCATTTATCATGTG 2952  
Db 2705 AAAGGAGCACAAAGCCATATCATGCTGTTGCTCCCATCTGCTACGACACATCTGGCAGCTG 2764

## RESULT 11

US-11-218-305-19385  
; Sequence 19385, Application US/11218305  
; Publication No. US2006014195A1  
; GENERAL INFORMATION:  
; APPLICANT: MONSANTO TECHNOLOGY, LLC  
; APPLICANT: McLaird, Paul L.  
; APPLICANT: tao, Nengbing  
; APPLICANT: Wu, Kunsheng

; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
; FILE REFERENCE: 38-21 (53660)B  
; CURRENT APPLICATION NUMBER: US/11/218,305  
; CURRENT FILING DATE: 2005-09-01  
; PRIOR APPLICATION NUMBER: US 60/606,880  
; NUMBER OF SEQ ID NOS: 25043  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 19385  
; LENGTH: 1710  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-11-218-305-19385

Query Match 1.6%; Score 52; DB 7; Length 1710;  
Best Local Similarity 51.3%; Pred. No. 0.009;  
Matches 121; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
Qy 2713 CATCCGGAACAACCTGTGGTAAACTTATCGTTTCGAAATACAAATTCGATTTTCTTGG 2772  
Db 1040 CAGCCGGCACTGTTGTGACACAAAGTCTGTCAATCCAGGAACTTCGATTTCTACATGT 1099  
Qy 2773 CATCTCATCATGTCCTTGGTACATCTCGTCCAGGACATTACACTGTTATGTATGACG 2832  
Db 1100 GTTCACATGCTGGAATGATTGGAACCTACCAGGCCAACTCACTATCATCTCTGCATGACG 1159  
Qy 2833 ATAAAGGAAATGAGCCAGATGAAGTCTATAAAATGACCTACGAGACTGCTTTCTCTCTG 2892  
Db 1160 AGTAGGCTTCAATCCTGATGACCTCGAGAGCTGGTGCACTGCTCTCTTACGTGTACC 1219  
Qy 2893 CTAGATGTCGAAACCCATCTCGTTGCTTCCGGTTCATTATGCTCATTTATCA 2948  
Db 1220 AAAGGAGCACAAAGCCATATCATGTTGTTGCTCCCATCTGCTACGCCACCTTGCA 1275

## RESULT 12

US-10-953-349-4063  
; Sequence 4063, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 4063  
; LENGTH: 3474  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-4063

Query Match 1.6%; Score 51; DB 6; Length 3474;  
Best Local Similarity 50.2%; Pred. No. 0.022;  
Matches 126; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
Qy 2698 GAATGTGAAACCCATCATCCGGAACAACTGTGATGAACCTATCGTTTCGAAATACAAAT 2757  
Db 2861 GAAGTGGGAATATTTTACCTGGCACTGTTGTGGACTCTAAATCTGCCACCTTACAGAGT 2920  
Qy 2758 TCGATTTTCTTGGCATCTCATCATGCTCTTGGTGTCTCTTGGTATCATCTCGTCAGGACATTACA 2817  
Db 2921 TTGACTTTTACCTCTGTAGTCAATGCTGGTATTTCAGGGCACTTCTCGACCTGCTCATACC 2980  
Qy 2818 CTGTTATGTATGACGATAAGGAATGAGCCAAAGATGAAGTCTATAAAATGACCTACGGAC 2877  
Db 2981 ACGTTCTTTGGGATGAGAACAACTTTACTGCAGATGGACTTCAATCTCTGACCAATACT 3040  
Qy 2878 TTGCTTTTCTCTGCTAGATGTCGAAACCCATCTCGTTGCTGCTGTTCCGGTTCATTATG 2937

Db 3041 TATGTTACAGTATGCAAGATGACACAGCGCTCAGTTTCAATTGTTCCCTCGCATATTATG 3100  
 QY 2938 CTCATTTATCA 2948  
 |||||  
 Db 3101 CACATCTAGCA 3111

RESULT 13  
 US-10-449-902-27837  
 ; Sequence 27837, Application US/10449902  
 ; Publication No. US20060123505A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: National Institute of Agrobiological Sciences.  
 ; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
 ; APPLICANT: The Institute of Physical and Chemical Research.  
 ; APPLICANT: Foundation for Advancement of International Research.  
 ; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
 ; FILE REFERENCE: MOA-A0205Y1-US  
 ; CURRENT APPLICATION NUMBER: US/10/449,902  
 ; PRIOR FILING DATE: 2003-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2002-203269  
 ; PRIOR FILING DATE: 2002-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2002-383870  
 ; PRIOR FILING DATE: 2002-12-11  
 ; NUMBER OF SEQ ID NOS: 56791  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 27837  
 ; LENGTH: 3432  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; PUBLICATION INFORMATION:  
 ; DATABASE ACCESSION NUMBER: AK103279  
 ; DATABASE ENTRY DATE: 2002-08-28  
 ; US-10-449-902-27837

Query Match 1.6%; Score 50.2; DB 6; Length 3432;  
 Best Local Similarity 50.2%; Pred. No. 0.034;  
 Matches 124; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
 QY 2698 GAATTGTGAACCATCATCCGGAACAACTGTGGATAACTTATCGTTTGGAAATACAAAT 2757  
 Db 2743 GGAGCGGGAACATATCTCCCTGGTACGTTGTAGATTCAAGATTCCGTCATCCAACTGAGT 2802  
 QY 2758 TCGATTTTCTTGGCATCTCATCATGTGCTCTTGGTACATCTCGTCCAGGACATTACA 2817  
 Db 2803 TTGACTTCTACCTGTGTAGCCATGCGGCATTAAGGGTACTAGTCGTCCAGCTCATTATC 2862  
 QY 2818 CTGTTATGTATGACGATAAGGAATGAGCCAAAGATGAAGTCTATAAAATGACCTACGGAC 2877  
 Db 2863 ATGCTCTGTGGATGAACAACTTCACAGCTGATGCGATTGCGAGATTCTTACCAACACC 2922  
 QY 2878 TTGCTTTTCTCTGCTAGATGTCGAAACCCATCTCGTTGCTGTTCGGTTTCATTATG 2937  
 Db 2923 TTGCTACACCTATGCAAGGTGCACTCGCTCTGTATCAATTGTTCCACCTGCTTATTATG 2982  
 QY 2938 CTCATTT 2944  
 |||||  
 Db 2983 CTCATCT 2989

RESULT 14  
 US-11-218-305-22488  
 ; Sequence 22488, Application US/11218305  
 ; Publication No. US20060141495A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MONSANTO TECHNOLOGY, LLC  
 ; APPLICANT: Monsanto  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Wu, Kunsheng  
 ; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
 ; FILE REFERENCE: 38-21 (53660)B  
 ; CURRENT APPLICATION NUMBER: US/11/218,305

; CURRENT FILING DATE: 2005-09-01  
 ; PRIOR APPLICATION NUMBER: US 60/606,880  
 ; PRIOR FILING DATE: 2004-09-01  
 ; NUMBER OF SEQ ID NOS: 25043  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 22488  
 ; LENGTH: 964  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: [941]..(942)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: [945]..(950)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: [952]..(953)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: [963]..(964)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; US-11-218-305-22488

Query Match 1.4%; Score 46.6; DB 7; Length 964;  
 Best Local Similarity 49.0%; Pred. No. 0.15;  
 Matches 124; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
 QY 2696 TGGATTGTGAACCCATCATCCGGAACAACTGTGGATAACTTATCGTTTGGAAATACAA 2755  
 Db 192 TAGAATGGCAACATATCTCCCGGCACTGTGTGGTGAAGTTTGCCTATCCCAACCA 251  
 QY 2756 ATTCATTTTCTTGGCATCTCATCATGTGTCTTGGTACATCTCGTCCAGGACATTATTA 2815  
 Db 252 GTTTCATTTTACCTGTGCGGCATGTGGCATTGAGGAAACAAAGCGTCTGCCCATTA 311  
 QY 2816 CACTGTTATGTATGACGATAAGGAATGAGCCAAAGATGAAGTCTATAAAATGACCTACCG 2875  
 Db 312 TCATCTCTGTGGATGAGAACAAATTTACGGCTGATGGGTTGCAAACTCTCACCACAA 371  
 QY 2876 ACTTGCTTTCTCTCTGTAGATGCGAAACCCATCTCGTTGCTGTTCGGTTTCATTATTA 2935  
 Db 372 CTTGTGTTACAGTATGCGAGTGACAGCTCAGTATCAATCGTTCTCTGCTGCTACTA 431  
 QY 2936 TGCTCATTTATCA 2948  
 |||||  
 Db 432 TGCTCATCTGGCA 444

RESULT 15  
 US-11-218-305-22485  
 ; Sequence 22485, Application US/11218305  
 ; Publication No. US20060141495A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MONSANTO TECHNOLOGY, LLC  
 ; APPLICANT: Monsanto  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Wu, Kunsheng  
 ; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping  
 ; FILE REFERENCE: 38-21 (53660)B  
 ; CURRENT APPLICATION NUMBER: US/11/218,305  
 ; CURRENT FILING DATE: 2005-09-01  
 ; PRIOR APPLICATION NUMBER: US 60/606,880  
 ; PRIOR FILING DATE: 2004-09-01  
 ; NUMBER OF SEQ ID NOS: 25043  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 22485  
 ; LENGTH: 965  
 ; TYPE: DNA

! ORGANISM: Zea mays  
US-11-218-305-22485

Query Match 1.4%; Score 46.6; DB 7; Length 965;  
Best Local Similarity 49.0%; Pred. No. 0.15;  
Matches 124; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy	2696	TGGAATTGTGAACCCATCATCCGGAACAACCTGTGGATAAATCTTATCGTTTCGAAATACAA	2755
Db	261	TAGAAGTGGCAACATACCTGCCGGGCACTGTGGGACTCGAAGATTGGCCATCCAACCGA	320
Qy	2756	ATTCGATTTTCTTTGGCATCTCATCATGTGTCTTGGTACATCTCGTCCAGGACATTA	2815
Db	321	GTTCGATTTCTACCTGTGCACCATGTGGCATTGAGGAACAAGCGTCTGCCCCATTA	380
Qy	2816	CACGTGTATGTATGACGATTAAGGAATGAGCCAAAGATGAAGTCTATAAAATGACCTACGG	2875
Db	381	TCATGTTCTGTGGGATGAGAACAAATTTACGGCTGATGGGTTGCAAACTCTCACCAACAA	440
Qy	2876	ACTTGCCTTTCTCTCTAGATGTCGAAACCCATCTCGTTGCCTGTTCCGGTTTCAATTA	2935
Db	441	CTTGTGTATACAGTATGCCAGGTGCACCGTCAAGTATCAATCGTTCTCTCTGCATACTA	500
Qy	2936	TGCTCATTATCA	2948
Db	501	TGCTCATCTGGCA	513

Search completed: July 5, 2006, 23:20:24  
Job time : 386 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 13:24:51 ; Search time 197 Seconds  
(without alignments)  
2367.313 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 5349  
Sequence: 1 MGSNPPEKGFYRHSLDPE.....RHMEHFLQTNKYPGMSFA 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq.8:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003ba:\*  
8: Geneseq2004s:\*  
9: Geneseq2005s:\*  
10: Geneseq2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5349	100.0	1020	AAU01856	Aau01856 C. elegans
2	604.5	11.3	997	ADQ80699	Adq80699 Arabidops
3	577	10.8	857	ADB81572	Adb81572 Human euk
4	577	10.8	857	ADJ94697	Adj94697 Human euk
5	577	10.8	857	AEC01709	Aec01709 Human EIP
6	577	10.8	857	AER73789	Aer73789 Human Arg
7	575.5	10.8	969	ABO43145	Abo43145 A. thalia
8	575.5	10.8	969	ADC46703	Adc46703 Thalecres
9	575.5	10.8	969	ADI43621	Adi43621 Plant tra
10	575.5	10.8	969	ADO02339	Ado02339 Thalecres
11	566	10.6	859	ABP64718	Abp64718 Human pro
12	566	10.6	859	AER73791	Aer73791 Human Arg
13	566	10.6	860	ABG97473	Abg97473 Human NAA
14	566	10.6	860	ADJ94698	Adj94698 Human euk
15	560.5	10.5	950	ABB67046	Abb67046 Drosophil
16	560.5	10.5	984	ABB62573	Abb62573 Drosophil
17	560.5	10.5	984	ABB67045	Abb67045 Drosophil
18	560.5	10.5	984	ABG70016	Abg70016 Larval vl
19	555.5	10.4	855	ADJ94700	Adj94700 Human euk
20	555.5	10.4	860	AAM80292	Aam80292 Human pro
21	555.5	10.4	860	AAB94291	Aab94291 Human pro
22	555.5	10.4	860	ABG72612	Abg72612 Human cyt
23	555.5	10.4	860	ADH77275	Adh77275 Human FAZ

24	555.5	10.4	860	10	AER73790	Aer73790 Human Arg
25	555	10.4	861	10	AER73792	Aer73792 Human Arg
26	555	10.4	924	7	ADJ69189	Adj69189 Human hea
27	555	10.4	924	8	ADJ94699	Adj94699 Human euk
28	554.5	10.4	860	6	ABU11762	Abu11762 Human MDD
29	550.5	10.3	858	9	ADM42496	Adm42496 Human euk
30	549.5	10.3	816	6	AAE37044	Aae37044 Human nuc
31	542.5	10.1	782	4	AAB93139	Aab93139 Human pro
32	529.5	9.9	585	8	ADP55696	Adp55696 Human PRO
33	523.5	9.8	990	8	ADO61721	Ado61721 Transcrip
34	523.5	9.8	990	8	ADQ80705	Adq80705 Arabidops
35	519	9.7	1048	3	AAG42118	Aag42118 Arabidops
36	519	9.7	1048	6	ADB17512	Adb17512 A. thalia
37	519	9.7	1048	8	ADR43623	Adr43623 Plant tra
38	519	9.7	1048	8	ADQ80695	Adq80695 Arabidops
39	519	9.7	1048	9	AEC75736	Aec75736 Arabidops
40	515	9.6	868	3	AAG42119	Aag42119 Arabidops
41	515	9.6	924	9	ADY65088	Ady65088 S. mansoni
42	510.5	9.5	1013	8	ADQ80701	Adq80701 Arabidops
43	509	9.5	1064	3	AAG52593	Aag52593 Arabidops
44	505	9.4	884	3	AAG52594	Aag52594 Arabidops
45	501	9.4	906	3	AAG51563	Aag51563 Arabidops

#### ALIGNMENTS

RESULT 1  
AAU01856  
ID AAU01856 standard; protein; 1020 AA.  
XX  
AC AAU01856;  
XX  
DT 07-SRP-2001 (First entry)  
XX  
DE C. elegans RNA interference pathway protein RDE-1.  
XX  
KW RNA interference; RNAi; RDE-1; genetic interference; antibody; dsRNA;  
KW double-stranded RNA-dependent gene silencing.  
XX  
OS Caenorhabditis elegans.  
XX  
PN WO200129058-A1.  
XX  
PD 26-APR-2001.  
XX  
PF 13-OCT-2000; 2000WO-US028470.  
XX  
PR 15-OCT-1999; 99US-0159776P.  
PR 30-MAR-2000; 2000US-0193218P.  
XX  
(UYMA-) UNIV MASSACHUSETTS.  
(CARN-) CARNEGIE INST WASHINGTON.  
XX  
Mello CC, Fire A, Tabara H, Grishok A,  
WPI; 2001-316239/33.  
N-PSDB; AAS03282, AAS03283.  
XX  
Novel RNA interference pathway genes and their protein products involved  
in mediation of genetic interference, useful for modulating and studying  
regulation of RNA interference pathway.  
XX  
Claim 3; Fig 6; 76pp; English.

The sequence represents the RNA interference (RNAi) pathway protein RDE-1. RDE-1 and RDE-4 are involved in the pathway mediating double-stranded RNA-dependent gene silencing (genetic interference) RDS-1 and RDS-4 protein is useful for preparing an RNAi agent, by incubating a dsRNA in the presence of the proteins. The prepared RNAi agents can be used as sequence-specific interfering agents for targeted genetic interferences. The nucleic acids are useful for studying the regulation of RNAi pathway and to generate knockout strains of animals such as C.elegans. RDE-1 and

CC RDE-4 genes and their products are useful for modulating RNAi pathway  
CC activity. The polypeptides are useful for generating and testing  
CC antibodies specific for the polypeptides which are useful for studying  
CC the RNAi pathway in C.elegans and other organisms. RNAi pathway genes are  
CC useful for mediating specific processes, e.g. a gene that mediates dsRNA  
CC uptake by the cells is useful for transporting other RNAs into cells or  
CC for facilitating entry of agents such as drugs into cells. RNAi pathway  
CC mutant strains (rde-1, rde-4 mutants) are useful in genetic screens to  
CC identify additional RNAi pathway components  
XX Sequence 1020 AA;

Query Match 100.0%; Score 5349; DB 4; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNFPPELEKGFYRHSLSLDPPEMKWLPARTGKCDGKFYEKKVLLVNMFKFSSKIYDREYVE 60  
DB 1 MSSNFPPELEKGFYRHSLSLDPPEMKWLPARTGKCDGKFYEKKVLLVNMFKFSSKIYDREYVE 60

QY 61 YEVKMTKEVLRNPKGPPPKTEIPIDRAKLFQWHLRHEKKQTFILEDYVDESKDTYV 120  
DB 61 YEVKMTKEVLRNPKGPPPKTEIPIDRAKLFQWHLRHEKKQTFILEDYVDESKDTYV 120

QY 121 SYCRLNTVTSKMLVSEKVKVKKSEKKDEKDLKKILYTMILTYRKKFHLNFSRENPEKDE 180  
DB 121 SYCRLNTVTSKMLVSEKVKVKKSEKKDEKDLKKILYTMILTYRKKFHLNFSRENPEKDE 180

QY 181 EANSYKFKLVNMTQKRYAPVNEBEIKVQFARNFYDNNISILRVPEFDPNRFQESLE 240  
DB 181 EANSYKFKLVNMTQKRYAPVNEBEIKVQFARNFYDNNISILRVPEFDPNRFQESLE 240

QY 241 VAPRIEAFGIYIGIKELPDGEPVLNFAIVDLKFNAPQMSLLDYLLIVDPQSCNDDVR 300  
DB 241 VAPRIEAFGIYIGIKELPDGEPVLNFAIVDLKFNAPQMSLLDYLLIVDPQSCNDDVR 300

QY 301 KDLTKLMAGKMTIQOAPRIRQLLENLKLKCAEVDNEMSRLETERHLTFDLCEENSL 360  
DB 301 KDLTKLMAGKMTIQOAPRIRQLLENLKLKCAEVDNEMSRLETERHLTFDLCEENSL 360

QY 361 VYKVTGSDRGNAKKYDTTLFKIYEENKKFTEFPHLPLVKYKGAKEVAVPMEHLEVHE 420  
DB 361 VYKVTGSDRGNAKKYDTTLFKIYEENKKFTEFPHLPLVKYKGAKEVAVPMEHLEVHE 420

QY 421 KPQRYKNRIDLVMQDKFLKRAIRKPHDYKENTLKMELDPSEBELNPFVERGLCSKLOM 480  
DB 421 KPQRYKNRIDLVMQDKFLKRAIRKPHDYKENTLKMELDPSEBELNPFVERGLCSKLOM 480

QY 481 IECPGKVLKEPMLVNSVNEIOIKMTPIRGFQEKQLNVVPEKELCCAVFVNNETAGNPCLE 540  
DB 481 IECPGKVLKEPMLVNSVNEIOIKMTPIRGFQEKQLNVVPEKELCCAVFVNNETAGNPCLE 540

QY 541 ENDVVKFYTELIGGCKFRGIRIGANENRGAQIMYDATKNEYAFYKNCITLNTGIGRFEIA 600  
DB 541 ENDVVKFYTELIGGCKFRGIRIGANENRGAQIMYDATKNEYAFYKNCITLNTGIGRFEIA 600

QY 601 ATEAKNMFERLPDQKQKLMFIIISKRLNAYGFVKYCHDHTIGVANQHITSETYTKALA 660  
DB 601 ATEAKNMFERLPDQKQKLMFIIISKRLNAYGFVKYCHDHTIGVANQHITSETYTKALA 660

QY 661 SURHEKSGKRIFYQALKNALKGINQELDWSAEISPEEKERRKTMPLTMYGIDVT 720  
DB 661 SURHEKSGKRIFYQALKNALKGINQELDWSAEISPEEKERRKTMPLTMYGIDVT 720

QY 721 HPTYSYSGIDYSIAAVASINPGGTIYRNMIIVTQEECRPGERAVAHGRERTDILEAKFVKL 780  
DB 721 HPTYSYSGIDYSIAAVASINPGGTIYRNMIIVTQEECRPGERAVAHGRERTDILEAKFVKL 780

QY 781 LREFAENNDRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKSVKQFMSEDEDEPEPKYT 840  
DB 781 LREFAENNDRAPAHIVVYRDGVSDSEMLRVSHDELRSLSKSVKQFMSEDEDEPEPKYT 840

QY 841 FVIQKRNHNTLLRRMEKDKPVVVKDLTPAETDVAVAVKQWEEDMKESKETGIYNPSSG 900

DB 841 FVIQKRNHNTLLRRMEKDKPVVVKDLTPAETDVAVAVKQWEEDMKESKETGIYNPSSG 900  
QY 901 TTVDKLIIVSKYKFDPFLLASHHGVLGTSRPGHYTYVMYDDKGMSSQDEYVYKMTYGLAFUSARC 960  
DB 901 TTVDKLIIVSKYKFDPFLLASHHGVLGTSRPGHYTYVMYDDKGMSSQDEYVYKMTYGLAFUSARC 960  
QY 961 RKPISLPVPVHYAHLSCCEKAKELYTYKHYHIGDYAQRTRHEMEHFLQTNVKYPCWMSFA 1020  
DB 961 RKPISLPVPVHYAHLSCCEKAKELYTYKHYHIGDYAQRTRHEMEHFLQTNVKYPCWMSFA 1020

RESULT 2  
ADQ80699  
ID ADQ80699 standard; protein; 997 AA.  
XX  
AC ADQ80699;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Arabidopsis thaliana TFL1-binding protein #7.  
XX  
KW TFL1-binding protein; plant growth control; biotechnology;  
KW fishing industry; screening.  
XX  
OS Arabidopsis thaliana.  
XX  
PN JP2004208572-A.  
XX  
PD 29-JUL-2004.  
XX  
PF 27-DEC-2002; 2002JP-00381220.  
XX  
PR 27-DEC-2002; 2002JP-00381220.  
XX (DOKU-) DOKURITSU GYOSEI HOJIN KAGAKU GIJUTSU SH.  
XX  
XX WPI; 2004-538191/52.  
DR N-PSDB; ADQ80698.  
XX  
PT Novel AGO protein, comprising PA2 and Piwi domain, specifically binding  
PT to TFL1 protein that is derived from Arabidopsis thaliana, useful for  
PT controlling growth phase of plant.  
XX  
PS Disclosure; SEQ ID NO 14; 104pp; Japanese.  
XX  
CC The invention comprises the amino acid and coding sequences of  
CC Arabidopsis thaliana proteins which specifically bind to TFL1 protein.  
CC The amino acid and coding sequences of the invention are useful for  
CC screening TFL1 family proteins. The DNA and protein sequences of the  
CC invention are also useful for controlling the growth phase of a plant,  
CC and as a reagent for research in biotechnology and fishing industry. The  
CC present amino acid sequence represents an Arabidopsis thaliana TFL1-  
CC binding protein of the invention.

Sequence 997 AA;  
Query Match 11.3%; Score 604.5; DB 8; Length 997;  
Best Local Similarity 23.8%; Pred. No. 1.8e-41;  
Matches 242; Conservative 164; Mismatches 353; Indels 259; Gaps 41;

QY 38 KKVLLLVNVPFSSKIYDREYVEYEVKMTKPKGNLNRKPKGPPKTEIPIDRAKLFWOHL 97  
DB 158 KKVVRAN--HFLVQVADRDLHYDVSNPEVSK----- 190

QY 98 RHEKKQTDFILEDYVDESKDTVYSYCRLNTVTSKMLVSEKVKVKKDSE---KKDEKDLKK 154  
DB 191 -----TVNRNVNMLLVKN---YKDSHLGGKSPAYD-GRK 220

QY 155 ILYTM--ILTYRKKFHLNFSRENPEKDEANRSYKF-LKNV-----MTQKVRVA 200  
DB 221 SLYTAGPLFPFSKEFVNVNLAEKRADGSGGKORPFAVKNVTSVDLYQLQQLDRKQREA 280



QY 201 PFVNEEIKVQPAKVFVYDNNISILRVPSFHDNRFEOSL-----EVAPRIEAMFGI 251  
 DB 281 PY--DTQVLVDVLRDPSNDYVSGRSP-----FHTSLGKARDGEGLDGLEYWRGY 333  
 QY 252 YIGIKELFDGEPVLFNAIVDKLFFYNAPKMSLLDYLIVDPQSCNDVVRKDKTKLMAGK 311  
 DB 334 FOSLR-LTOMGLSLNIDVSARSFY-----EPVVTDPIKFLNIRDL--N 375  
 QY 312 MTIQAARPRIRQLLENLKLKCAVWDMNSRLTERHLTFLDLCEENSLVYKTKGSD-- 369  
 DB 376 RPLRDSRLKVKVRLTKVKKLH-WN-----CTKSA---KISSISLP 415  
 QY 370 ---RGRNAKYDTTLFKIEENKCF-LEPHPLPLVKVKGAKAYAVPMHELVHEKPOR 424  
 DB 416 IRELFTLEDKSEKTVVQFAEKYKRVQALPAITGSDTRPVYLPMELCQIDE-QOR 474  
 QY 425 YKRNIDVMDQFLKRAKTRPHDYKENTLKMELDFSSBELNFVERFGLCSKLMTECP 484  
 DB 475 YTKRLNEKQVALLKATQRPDP--RENSIKNLVKNYNDLS--KEFGMSVTTQLASIE 531  
 QY 485 GKVLKEPLMNVSNQIQTMPVIRG-----FOEKQLNVVPEKELCCAVFVNVETAGNPCLE 540  
 DB 532 ARVLPPEMLKYHDSGKEMVNPRLGWNMDKKMVGAKVTSWTCVSFSTRIDRGLP--- 588  
 QY 541 ENDVVKFYTELIGCKFERGIRIGANENRGAQSIWYDATKNEYAFYKNTLNTVIGREPIA 600  
 DB 589 ---QBFCKGLIGVCVSGM-----EPKQPAIPFISCPPEH-----IEEA 625  
 QY 601 ATEAKMFERLPDKEQKVLFFIISKRQLNAYGFVHKYCHDHTIGVANOHTSETVTKALA 660  
 DB 626 LLD---IHKEAPGLQ---LLVILPDVTGSGYKIKKICETELGIYSQCCQPRQVNL-- 676  
 QY 661 SLRHEKSGRIFFQIAKINAKLGGINQELDWSIEAISPKEKERRKTMPL-----TMYV 715  
 DB 677 ---NKOYMENVALKINVTGGRNTVLN-----DAIRRNIPILITDRPTIIM 718  
 QY 716 GIDVHTPTSYSIDYSTAAVASIN-PGGTIYRNMIVTQ---EE-----CEPGERAV 763  
 DB 719 GADVTIHPQGEDSSPSAAVASMDWPEINRYKGLVSAQAHRSEITODLYKLVQDPORGL 778  
 QY 764 AHGRERTDILEAKFVLLRFAENNDRAPAHVTVYRDGVSDSEMLRVSHDELSLSEV 823  
 DB 779 VH-----SGLIREHPIAFRA-----TCQIPQRIIFYRDGVSEGFQSVQLHEMTAKAC 829  
 QY 824 KQFMSERDGEDPPEKTYFVIQKRNHTLLRRMEKPKVNVNOLDTPAETDVAVAVKQWE 883  
 DB 830 NSLQ-----ENYVPRVTVIVQKRRHTRLPPEQHG-----NRDWT----- 864  
 QY 884 EDMKESKETGVNPSSTGTTVDKLVSKYKDFDFFLASHGVGLTSRPGHYVTMTDDKMSQ 943  
 DB 865 -----DKSGNIQF--GTVVDTKICHNPNEPFDYLNHSHAGIQTSRPAHYHVLDDENGFTA 916  
 QY 944 DEVYKMTYGLAFLSARCKRPISLPVPHVAHLSCEKAKELYRYTKHYIGDYAQPRT 1001  
 DB 917 DQLQMLTNLNLCTYVARTKSVISVPPAYAHAAFR-----RYMSEMSDGGSSRSR 970

RESULT 3  
 ADB81572  
 ID ADB81572 standard; protein; 857 AA.  
 XX  
 AC ADB81572;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human eukaryotic translation initiation factor 2C 1 protein.  
 DE  
 XX human; eukaryotic translation initiation factor 2C 1; EIF2C1;  
 KW chromosome 1p34-35; Co-EIF2C; EIF2C; Golgi ER protein 95kDa; GERP95; Q99;  
 KW gene therapy; hyperproliferative disorder;  
 KW familial hypercholesterolemia; cancer; polycystic kidney disease;  
 KW cystic fibrosis; progeria syndrome; cytostatic; antileukemic.

OS Homo sapiens.  
 XX WO2003040321-A2.  
 XX  
 XX 15-MAY-2003.  
 XX  
 XX 04-NOV-2002; 2002WO-US035324.  
 XX  
 XX 08-NOV-2001; 2001US-00007078.  
 XX  
 XX (ISIS-) ISIS PHARM INC.  
 XX  
 XX Ward DT, Watt AT;  
 XX  
 XX WPI; 2003-449448/42.  
 XX  
 XX N-PSDB; ADB81486, ADB81493.  
 XX  
 XX New compound, having a sequence targeted to a nucleic acid encoding human  
 PT collapsin response mediator protein 2, useful for preparing a composition  
 PT for treating hypercholesterolemia or hyperproliferative disorder, e.g.,  
 PT cancer.  
 XX  
 XX Disclosure; Page 82-86; 120pp; English.  
 XX  
 CC This invention relates to novel antisense oligonucleotides that modulate  
 CC the expression of human eukaryotic translation initiation factor 2C 1  
 CC (EIF2C1). EIF2C1 is located on chromosome 1p34-35, and is also known as  
 CC Co-EIF2C, EIF2C, Golgi ER protein 95kDa, GERP95 and Q99. It is an  
 CC intracellular membrane associated protein thought to be involved in  
 CC cellular differentiation, such that altered expression of EIF2C1 can  
 CC affect cell growth, morphology and tumorigenicity. Accordingly,  
 CC antisense oligonucleotides that inhibit the expression of EIF2C1 in cells  
 CC or tissues can be used in gene therapy to treat various conditions  
 CC including hyperproliferative disorders, familial hypercholesterolemia  
 CC and cancer, as well as polycystic kidney disease, cystic fibrosis and  
 CC progeria syndrome. As such, the oligos of the present invention can be  
 CC described as having cytostatic and antileukemic activities. This  
 CC polypeptide sequence is the human eukaryotic translation initiation  
 CC factor 2C 1 (EIF2C1) protein of the invention. NOTE: This protein  
 CC sequence is not further described in the specification but is given as an  
 CC embedded protein in the sequence listing.  
 XX  
 XX Sequence 857 AA;  
 SQ

Query Match 10.8%; Score 577; DB 7; Length 857;  
 Best Local Similarity 23.0%; Pred. No. 3.1e-39;  
 Matches 227; Conservative 171; Mismatches 365; Indels 222; Gaps 37;

QY 38 KKVLVLLVNNPKFSKSIYDREYVEYVMTKEVNLKPKGPPKTEIPIPRAKLFWOHL 97  
 DB 34 KPIKLLANYPEVDIPKIDVYHVEVDIKPKD-----CPRVNVREVVE-----YVW 77  
 QY 98 RHEKKTQDTILEDY--VFDEKDTVYVCRNLNTVTSKMLVSEKV---VKDSEKDEKDL 152  
 DB 78 QHFKPQ---IFGDRKPVYDGKNYI-----VTALPIGNERVDFVTIPEGKD-RIFK 127  
 QY 153 KKILYTMILTIRKKFHLNFSRENPEKDEANRSYKFLKNVMTQKRVYAPVNEIKVQFA 212  
 DB 128 VSIKWLAIYSWRMLHEALVSGQIPVPLESVQALDVARHLAS--MRYTPV----- 175  
 QY 213 KNFVYDNNISILRVPSFHDNRFEOSLVAPRIEAMFGIYIGIKELPDGSPVLFNAIVDK 272  
 DB 176 -----GRSFFSPPEGYTHP-----LGGREVVNFGFQSVRPAM-WQMLNIDVSAT 220  
 QY 273 LFVNAPKMSLLDYLIVDPQSCNDVVRKDKTKLMAGKMTIRQAARPRIRQLLENLKLK 332  
 DB 221 AFYKA--QPVIEFMCVSLDIRNIDQPK-----PLTDSQVRFTKIGKLAKE 266  
 QY 333 ---CAVWDN---EMSLTERHLTFLDLCEENSLVYKTKGSDRGNACKYDTTLFKYI 385  
 DB 267 VTHCGMKKRYKRVNTRRPASHQTFPLQLSQTV-----ECTVAQYF 310  
 QY 386 EENKCF-IEFPHPLPLVKVKGAKAYAVPMHELVHEKPORYKRNIDLVMDQKFLKRAK 444

Db 311 KKYNNQLKYPHLPCLQVQEQKHLYPLEVNCI--VAGQRCIKKLTNDNOTSTMIKATARS 369  
QY 445 PHDYKENTLUMKELDFSSSEELNVERFGLCSKQLMIECPGKVLKEPMLVNSVNEQIKWT 504  
Db 370 APDRQEEISRLMKNASYNLDP--YIQEFGIKVKDDMTVTGRVLPAPILQYGGNRRAIAT 427  
QY 505 P-----VIRGFEKQL--NVVPEKELCCAVPVNNEAGNCLNEENDVVKFYTELI----- 552  
Db 428 PNQGVMDMRG---KQFYNGIEIKWAIACFAPQK-----QCREE--VLKNFTDQLRKISKD 478  
QY 553 GCKFRGIRIGANENRGAQSIIMYDATKNEYAFYKNCI--VAGQRCIKKLTNDNOTSTMIKATARS 612  
Db 479 AGMPIQGGPCFKYAGGADSV-----EPHFRHLK 507  
QY 613 DKEQKVLMPFIISIKQLNAYGFVGHYCDHTIGVANHITSETVTYKALASLRHEKSKRIF 672  
Db 508 NTYSGQLIIVILPGKTPVYAEKRVGDTLLGMATQCVQVKNVKT-----SPQTL 558  
QY 673 YQIALKINAKLGGINQELDWSIAEISPEEKERRKTMPLTMVIGDVTHTPTSYSGIDYSI 732  
Db 559 SNLCLKINVKLGGINNIL-----VPHQSAVFQPPV--IFLGADVTHTPPAGDGKPKSI 609  
QY 733 AAVVASINPGGTIYRNMIVTQECRPGERAVAHGRERTDILE--AKFVK--LLREFAENND 789  
Db 610 TAVGSMDAHPSRYCATRVQ-----RPRQEIIEDLSYMWRELLIQFYKSTR 656  
QY 790 NPRAHIVVYRQVSDSEMLRVSHDELRSKSEVKQFMSERDGEDPEPKYTFIVYQKRN 849  
Db 657 FK--PTRIIFRQVGPQGLQILHVELLAIRDACIKL-----EKDYQPGITVIVYQKRN 710  
QY 850 TLLRRMEKDKPVNKKLTPTAETDAVAVAKVQWEDMKESKETGIVNPSGGTTVDKLIUS 909  
Db 711 TELF--CADKNERIGKSGNIPA-----GTTVDNTNITH 740  
QY 910 KYKFOFFFLASHHGVLGTSRPGHYTMVYDDKMSQDEVYKMTYGLAFLSARCKRPSLVP 969  
Db 741 PEPDFYLCSHAGIOGTSRPSHYVYVLDNDRRTADELQILTYQLCHYVYCRTSVSIPAP 800  
QY 970 VHYAHLSCBKAKELYRTYKEHYGD 994  
Db 801 AYYARLVAPRAR--YHLVDKEHDSGE 824

## RESULT 4

ADU94697  
ID ADJ94697 standard; protein; 857 AA.

AC ADJ94697;

DT 06-MAY-2004 (first entry)

DE Human eukaryotic translation initiation factor 2C1 (eIF2C1) protein.

KW cytostatic; antiinflammatory; virucide; immunosuppressive; tumour;

KW inflammatory; infectious disease; viral infection; degenerative;

KW autoimmune; gene therapy; argonauce family;

KW eukaryotic translation initiation factor 2C1; eIF2C1; human.

OS Homo sapiens.

XX WO2004007718-A2.

XX 22-JAN-2004.

XX 10-JUL-2003; 2003WO-BP007516.

XX 10-JUL-2002; 2002EP-00015532.

XX 23-AUG-2002; 2002EP-00018906.

XX (PLAC ) MAX PLANCK GBS FOERDERUNG WISSENSCHAFTEN.

XX Tuschl T, Martinez J, Patkaniowska A, Urlaub H, Luehrmann R;

XX WPI: 2004-122948/12.  
DR N-PSDB; ADU94703.

XX New single-stranded RNA molecule having a length from 14-50 nucleotides,  
PT useful for preventing or treating tumor, inflammatory, infectious, e.g.  
PT viral infections, degenerative and autoimmune diseases.

XX Example; Fig 14; 73pp; English.

XX The invention relates to a novel single-stranded RNA molecule having a  
CC length from 14-50 nucleotides where at least 14-20 of the 5'-most  
CC nucleotides are substantially complementary to a target transcript. The  
CC RNA molecule of the invention demonstrates cytostatic, antiinflammatory,  
CC virucide and immunosuppressive activities and may be useful for  
CC inhibiting the expression of a target gene in vitro or in vivo,  
CC preferably for preventing or treating diseases associated with the  
CC overexpression of at least one target transcript. The diseases may be  
CC selected from tumour diseases, inflammatory diseases, infectious diseases  
CC such as viral infections, degenerative diseases and autoimmune diseases.  
CC Furthermore, the molecules of the invention may be utilised during gene  
CC therapy. The current sequence is that of the human eukaryotic translation  
CC initiation factor 2C1 (eIF2C1) protein of the invention.

XX Sequence 857 AA;

QY Query Match 10.8%; Score 577; DB 8; Length 857;

QY Beat Local Similarity 23.0%; Pred. No. 3.1e-39;

QY Matches 227; Conservative 171; Mismatches 365; Indels 222; Gaps 37;

QY 38 KVVLLVNVKFSKSIYDREYVEYVQMTKEVLRNKPCKPKKTEIPIDRAKLPWQHL 97

QY 34 KPIKLLANFYFEVDIPKIDVHYEVDIKPKD-----CPRVNRVEVE-----YVW 77

QY 98 RHEKQTDFILEDY--VFDEKDTVYVSVCLNTVTSMVLSEKV---VKDSKSKDEKDE 152

QY 78 QHFKPQ---IFGDRKPVYDGKNIY-----TVTALPIGNERVDPEVTIPGEGKD-RIFK 127

QY 153 KILYTMILTTRKPHLNPFSRENPEKDEANRSYKFLKNVMTQKRYAPFVNEEIKVQPA 212

QY 128 VSIKWLAIWSRMLHEALVSGQIPVPLESVQALDVAHRHLAS--MRYTPV----- 175

QY 213 KNFVYDNNISILRVPSFHPNRFQESLEVPARIEAFGIYIGIKELFQCEPVLNFAIVDK 272

QY 176-----GRSFPSPGYYHP-----LGGREVFQGHQSVRPAW-KWMLNLDVSAT 220

QY 273 LFNAPKMSLLDYLLIYDPOSCNDDVRKDLTKLMAGQMTIRQAARPIROLLENLKK 332

QY 221 AFYKA--QPVEFMCEVLDIRNIDBQPK-----PLTDSQVRVFTKEIKGLKVE 266

QY 333 ---CAEVWDN---EMSLRTERHLTFLDLCENSLVYKVTGKSDRGNKKYDITLTKIY 385

QY 267 VTHCGQMKRYVNCVTRRPPASHQTFPLQLESGQTV-----ECTVAQYF 310

QY 386 BENKCF-IEFPHLPVLYKSGAKYAVPMEHLEVHEKQRYKNRIDLVNQDFLKRATR 444

QY 311 KQKNLQKLYPHLPCLQVQEQKHLYPLEVNCI--VAGQRCIKKLTNDNOTSTMIKATARS 369

QY 445 PHDYKENTLUMKELDFSSSEELNVERFGLCSKQLMIECPGKVLKEPMLVNSVNEQIKWT 504

QY 370 APDRQEEISRLMKNASYNLDP--YIQEFGIKVKDDMTVTGRVLPAPILQYGGNRRAIAT 427

QY 505 P-----VIRGFEKQL--NVVPEKELCCAVPVNNEAGNCLNEENDVVKFYTELI----- 552

QY 428 PNQGVMDMRG---KQFYNGIEIKWAIACFAPQK-----QCREE--VLKNFTDQLRKISKD 478

QY 553 GCKFRGIRIGANENRGAQSIIMYDATKNEYAFYKNCI--VAGQRCIKKLTNDNOTSTMIKATARS 612

QY 479 AGMPIQGGPCFKYAGGADSV-----EPHFRHLK 507

QY 613 DKEQKVLMPFIISIKQLNAYGFVGHYCDHTIGVANHITSETVTYKALASLRHEKSKRIF 672

QY 508 NTYSGQLIIVILPGKTPVYAEKRVGDTLLGMATQCVQVKNVKT-----SPQTL 558



**Qy** 970 VHYAHLSCAKELRYTYKEHYIGD 994  
:  
:  
:  
**Db** 801 AYYARLVAEFRAR-YHLYVDKEHDSGE 824

## RESULT 6

AEF73789  
ID AEF73789 standard; protein; 857 AA.

AC AFF73789;

DT 06-APR-2006 (first entry)

Human Argonaute protein hAgo1.

Argonaute; hAgo1; RNA interference; RNAi; drug screening;  
KW  
X-ray crystallography; gene silencing; protein structure;  
KW

OS Homo sapiens.

FT	Location/Qualifiers	230. .350
FH	Key	
FT	Domain	

FT	/label = PAZ domain
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

FT	Domain	581..830	1
FT		/label = PIWI domain	

FT /label = PIWI domain

PN WO2006015258-A2.

PD 09-FEB-2006.

28-JUL-2005: 2005WO-US027084.

28-JUL-2004: 2004US-0592297P.

PR 29-JUL-2004; 2004US-0592269P;

PA (COLD-) COLD SPRING HARBOR LAB.

PI Joshua-Tor L, Song J, Hannon GJ, Liu J, Carmell MA, Rivas F;  
PI Marsden C;

WPI: 2006-155768/16.

Crystalline Argonaute useful for identifying agent that binds Argonaute PT  
PT protein, utilized for treating pancreatic cancer, treating hepatitis C PT  
PT infection, cancer or inflammatory diseases. PT

PS Disclosure: SEQ ID NO 1; 215pp; English.

The present sequence is that of human Argonaute protein hAgo1. Argonaute proteins are involved in RNA interference (RNAi). The invention provides a crystalline Argonaute protein, which comprises (i) an N-terminus, middle and PIWI domain which form a crescent-shaped base, and (ii) a PAZ domain, which is positioned above the crescent-shaped base, resulting in a cleft between the crescent-shaped base and the PAZ domain. The structure of the full-length Argonaute protein AseF73751 from the archaeobacterium *Pyrococcus furiosus* (PfAgo), as determined by x-ray crystallography to 2.25 Angstrom resolution, is provided. The invention also provides an isolated complex comprising an Argonaute protein and a single-stranded RNA hybridized to its target nucleic acid. In certain embodiments, the single-stranded RNA is bound to the PAZ domain of the Argonaute protein, and may further interact with the crescent-shaped base of the Argonaute protein. Also claimed are: a method of determining the three-dimensional structure of an Argonaute protein or its mutant, derivative, variant, analog, homolog, sub-domain or fragment by alignment of the PfAgo amino acid sequence to match homologous regions; a method of identifying an agent that binds an Argonaute protein by applying a three-dimensional molecular modeling algorithm to the atomic coordinates of an Argonaute protein to determine the spatial coordinates of the binding pocket of the Argonaute protein, and electronically screening the stored spatial coordinates of a set of candidate agents against the spatial coordinates of the Argonaute protein binding pocket; a computer-based method for the analysis of the interaction of a molecular structure



```

Db      565  -----WVAGAK---VTSWTCFKPQPAIPFISCPPEH-----IEEALLD- 600
Qy      605  KMFELPDKQKQVLFIIISKRQLNAYGFVXHYCDHTTIGVANQHTTSTVTTKALASLRH 664
Db      601  --IHKAAPGLQ---LLVILPDVTGSGYKIKRICEITELGIVSQCCQPRQVKNL----- 648
Qy      665  EKGSKRIFQVIALKINAKLGGGINQELDWSEIAEISPEEKERKTMPL-----TMYVGIDV 719
Db      649  ---NKOYMEVNAKINVKVGGRTVLN-----DAIRNRNPLITDRPTIINGADV 694
Qy      720  THPTSYSGIDYSIAAVASIN-PGGTIYRNMIWQ---EE-----CRPGERAVAHGR 767
Db      695  THPQGEDSSPIAAVAVASMDPNEINKYGLVSAQAHRREIIOQLYKLVQDPQRGVLH-- 752
Qy      768  ERTDILEAKFKVLLREFENNDRAPAHIVVYRDGVSDSEMLRVSHDELSLKSEVKQPM 827
Db      753  --SGLIREHFIAPRRA-----TQIQIPQRIIFYRDGVSEGFQSVLLHEMTAIRKACNSLQ 805
Qy      828  SERDGEDPEPKYTFVIOKRNHTLLRRMEKDKPVVVKDLTAEITDVAVAVKQHEEDMK 887
Db      806  -----ENYVPRVTFVIVQKHHHTLFPPEQHG-----NRDMT----- 836
Qy      888  ESKETGINVPSGTTVDKLIIVSKYKDFPLASHHGVLTGTSRPGHYVTWYDDKMGMSODEVY 947
Db      837  ---DKSGNIQP--GTVDVTKICHNEPFDYLNSHAGIQTSRPAHYVHLLDENGFTADQLQ 892
Qy      948  KWTYGLAFISARCKPISLPVPHYVLAHSCAKELRYTKHYIGDYAQPRTR 1001
Db      893  MLTNLCYTYARCTKSIVSVPPAYVAHLAAFR-----RYMSEMSDGGGSRSR 942

```

## RESULT 8

ADC46703  
ID ADC46703 standard; protein; 969 AA.

AC ADC46703;

DT 18-DEC-2003 (first entry)

DE Thalecress transcription factor-like protein G1149.

KW Thalecress; transcription factor-like protein; seed trait; transgenic;  
KW plant size; stress tolerance; yield; disease resistance; plant.

OS Arabidopsis thaliana.

XX US2003093837-A1.

PN 15-MAY-2003.

PF 01-NOV-2002; 2002US-00286264.

PR 23-MAR-1999; 99US-0125814P.

PR 22-MAR-2000; 2000US-00533030.

XX (KEDD/) KEDDIE J.

PA (RIEC/) RIECHMANN J L.

PA (RATC/) RATCLIFFE O.

PA (ZHAN/) ZHANG J.

PA (JIAN/) JIANG C.

PA (PINE/) PINEDA O.

PA (HEAR/) HEARD J.

PA (YUGG/) YU G.

PA (ADAM/) ADAM L.

PA (BROU/) BROUN P.

PA (REUB/) REUBER L.

PA (PILG/) PILGRIM M.

PA (SANA/) SAMAH R.

XX Keddle J, Riechmann J L, Ratcliffe O, Zhang J, Jiang C, Pineda O;

PI Heard J, Yu G, Adam L, Broun P, Reuber L, Pilgrim M, Samaha R;

XX

DR WPI: 2003-765498/72.  
DR N-PSDB; ADC46702.

XX Novel transgenic plant having recombinant polynucleotide encoding  
PT polypeptide that alters trait of transgenic plant when compared with same  
PT trait of another plant lacking recombinant polynucleotide.

XX Claim 1; SEQ ID NO 102; 165pp; English.

CC The invention relates to a transgenic plant having recombinant  
CC polynucleotide (II) encoding polypeptide comprising at least 6  
CC consecutive amino acids of a sequence chosen from the protein sequence  
CC appearing as ADC46603 - ADC46749 (every second sequence), where  
CC recombinant polynucleotide alters a trait of the seed transgenic plant  
CC when compared with same trait of another plant lacking recombinant  
CC polynucleotide. The proteins are transcription factor-like proteins. Also  
CC included are altering (M1) a trait associated with seed (comprising:  
CC transforming a plant with (II); selecting the transformed plants; and  
CC identifying a transformed plant with seed having altered trait), altering  
CC (M2) the expression levels of at least one gene of a plant (involving  
CC transforming the plant with (II) and selecting the transformed plant),  
CC altering (M3) a trait associated with a plant's seed (comprising:  
CC transforming the plant with a recombinant polynucleotide comprising a  
CC nucleotide sequence comprising least 18 consecutive nucleotides of a  
CC sequence appearing as ADC46750 - ADC46766 and selecting the transformed  
CC plan) altering (M4) a plant's trait (involving providing a database  
CC sequence, comparing the database sequence with a polypeptide or a  
CC polynucleotide chosen as detailed above, selecting a database sequence  
CC that needs selected sequence criteria and transforming a database  
CC sequence in the plant) and altering a plant's trait (involving providing  
CC a test polynucleotide, hybridising the test polynucleotide with a  
CC polynucleotide detailed above at low stringency and transforming the  
CC hybridising test polynucleotide in a plant to alter a trait of the  
CC plant). The method (M1) is useful for altering a trait associated with  
CC seed. The method (M2) is useful for altering the expressing levels of at  
CC least one gene of a plant. The method (M3) is useful for altering a trait  
CC associated with a plant's seed. The method (M4) is useful for altering a  
CC plant's trait. The method (M4) is useful for altering a plant's trait  
CC such as seed or plant size, stress tolerance, yield or disease  
CC resistance. The present sequence represents a transcription factor-like  
CC protein/seed trait altering protein of the invention.

XX Sequence 969 AA;

Query Match 10.8%; Score 575.5; DB 7; Length 969;

Best Local Similarity 23.2%; Pred. No. 56-39; Indels 279; Gaps 40;  
Matches 235; Conservative 161; Mismatches 339;

```

Qy      38  KKVLILLVNWKFSSKIYDREYVEYVQMTKEYLNRKPGKPFPPKTEIPIPRAKLFWOHL 97
Db      158  KKVWYRAN--HFLVQVADRDLHYDVSINPEISK----- 190
Qy      98  RHEKKQDTFILEDYVDEKDTVYSVCRNLNTVTSKMLVSEKVKVKKDSE---KKDEKLEKK 154
Db      191  -----TVNRNVMKLLVKN---YKDSHGLGKSPAYD-GRK 220
Qy      155  ILTYM--ILTYRKKFHLNFSRENPEKDEANRSYKF-LKNV-----MTQKRVYA 200
Db      221  SLYTAGLPFFDSKEFVNLAEKRGSSGKGKPFKVVAVKNTSTDLYQQLQFDRKQREA 280
Qy      201  PFVNEEIKVQFAKNFVYDNNISILRVPESPHDPNRFQSL-----EVAPRIEAWFGI 251
Db      281  FY--DTIQVLVDVLRDKPSNDYVSGRSF-----FHTSLGKDARDGSGELGDIYWRGY 333
Qy      252  YIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLILLIVDPSCNDVDKDLTKLMAKG 311
Db      334  FQSLR-LTQMGLSLNIDVSARSFY-----EPIVVTDFISKFLNIRDL--N 375
Qy      312  WTIROAARPRIQLLENLKLKCAEYVDNEMSLRTERHLTFLDLCEBSNLVYKVTGKSD-- 369
Db      376  RPLRSDRLKVKVLRUTLKVLLH-WNGTKSA-----KISGISLSP 415
Qy      370  ----RGRNAKKYDTTLTKIYEENKKF-IEFPHPLPLVKVSKGAKEAYAVPMHEHLEVBKQ 424

```



```
Db 416 IRLRPLEDKSEKTVQVFAKKNYKQALPAIQTGSDTRVYUPELMCLUCIDE-QOR 474
Qy 425 YKNRIDLVMQDFKRAKTRPHDYKENTLKMELKELDFSSBELNPFVERFGLCSKLMQIECP 484
Db 475 YTKRLNEKQVALLKATCORPPD-RENSINKLVKNYNDLS--KEFGMSVTTQLASIE 531
Qy 485 GKVLKEPMLVNSNEQIKMTPVIRGFQBEKQLNVVPEKELCCAVFVNNETAGNCPLENDV 544
Db 532 ARVLPPLPKYHDSGKGMNPRLG-----QWNMDKK----- 564
Qy 545 VKFYTELIGGCKFRGIRGANENRGAOSIMYDANKNEYAFYKNCNTLNTGIRPEIAATEA 604
Db 565 -----MVNGAK---VTSWCEPKPQPAIPFISCPPEH-----TEALLD- 600
Qy 605 KMFPERLPDKEQKVMFIIISKQLNAYGFVKYCHDHTIGVANQHITSETVTTKALASLRH 664
Db 601 --IHKRAPGLQ-----LLVILPDPVTGSGYKIKRICETELGIVSCCQPRQVNL----- 648
Qy 665 EKGSKRIFYQIALKINAKLGINQELDWSIAEISPEEKERRKTMPL-----TMYGIDV 719
Db 649 ---NKQTMENVALKINKVTCGRNTVLN-----DAIRRNIPLTDRPTIIMGADV 694
Qy 720 THPTSYSGIDYSIAAVASIN-PGTTIYRNMTVO---EE-----CRPGRAVAHGR 767
Db 695 THPOGEDSPSIAAVASMDWEINIKYRGLVSAQAUREEIIQDLYKLVQDPQORGLVH-- 752
Qy 768 ERTDILEAKFVKLLREPAENNDNRAPAHIVVYRDGSDSEMLRVSHDELRSKSEVKQFM 827
Db 753 --SGLIREHFAFRA-----TGQIPQRIIFYRDGVSEGOFSQVLLHEMTAIRKACNSLQ 805
Qy 828 SERGDGPEPKYFIVIOKHNTRLLRMKDKPVPVNNKDTTPAETDVAAVAKQWEEDMK 887
Db 806 -----ENVYPRVTVFVIOKRRHTRLPPEQHG-----NRDMT----- 836
Qy 888 ESKETGVNPSSTTVDKLVSKYKPDFFLASHGVLTGRPGHYTVMYDDKMGSDDEVY 947
Db 837 --DKSGNIQP--GTVDUTKICHNEFDYILNSHAGIQGTSRPAHYHVLLENGFTADQLQ 892
Qy 948 KMTYGLAFLSARCKPISLPVPVHYAHLSCAKELYRTYKHEYIGDYAQPRTR 1001
Db 893 MLTNNLCYTVARCTKSVSIVPVAYAHAAAFRA----RYTMESEMSDGGSSRSR 942

RESULT 9
ADI43621
ID ADI43621 standard; protein; 969 AA.
AC ADI43621;
XX
DT 22-APR-2004 (first entry)
XX
DE Plant transcription factor #683.
XX
KW transgenic; plant; enhanced tolerance to abiotic stress;
KW phosphate tolerance; hormone sensitivity; disease resistance;
KW sugar sensing; flowering; flower structure; stem bifurcation;
KW branching pattern; apical dominance; trichome; stem morphology;
KW root growth; root hair; seed development; cell proliferation;
KW cell differentiation; premature senescence; necrosis; plant size;
KW leaf morphology; seed morphology; seed biochemistry; root anthocyanin;
KW plant anthocyanin; light response; shade avoidance; bioinformatic;
KW transcription factor; gene; ds.
XX
OS Arabidopsis thaliana.
XX
XX US2004019927-A1.
XX
XX 29-JAN-2004.
XX
XX 25-FEB-2003; 2003US-00374780.
XX
XX 18-APR-2001; 2001US-00837944.
PR
```

```
XX (SHER/) SHERMAN B K.
PA (RIEC/) RIECHMANN J L.
PA (JIAN/) JIANG C.
PA (HEAR/) HEARD J E.
PA (HAAR/) HAAKE V.
PA (CREE/) CREELMAN R A.
PA (RATC/) RATCLIFFE O.
PA (ADAM/) ADAM L J.
PA (REUB/) REUBER T L.
PA (KEDD/) KEDDIE J.
PA (BROU/) BROUN P E.
PA (PILG/) PILGRIM M L.
PA (DUBE/) DUBELL A N.
PA (PINE/) PINEDA O.
PA (YUGG/) YU G.
XX
PI Sherman BK, Riechmann JL, Jiang C, Heard JE, Haake V,
PI Creelman RA, Ratcliffe O, Adam LJ, Reuber TL, Keddie J, Broun PE,
PI Pilgrim ML, Dubell AN, Pineda O, Yu G;
XX WPI; 2004-132245/13.
XX
XX New transgenic plant comprising a recombinant polynucleotide of any one
XX of more than 500 nucleotide sequences, useful in bioinformatic search
XX methods.
XX
PS Claim 1; SEQ ID NO 2084; 435pp; English.
XX
CC The invention describes a transgenic plant comprising a recombinant
CC polynucleotide of any one of more than 500 nucleotide sequences fully
CC defined in the specification or its complement. The method of the
CC invention can be used to produced a plant having altered traits such as:
CC enhanced tolerance to abiotic stress; glyphosphate tolerance; hormone
CC sensitivity; disease resistance; sugar sensing; early or late flowering;
CC altered flower structure, change in stem bifurcations, altered branching
CC pattern, reduced apical dominance, reduced trichome density; lack of
CC trichomes; increased ectopic trichome development; altered trichome
CC development; increase in trichome number; altered stem morphology;
CC increased root growth; increased root hairs; altered seed development;
CC altered cell proliferation or cell differentiation; rapid development;
CC premature senescence; increased necrosis; increase in seedling or plant
CC size; decreased plant size; leaf morphology; seed morphology; seed
CC biochemistry; increase in root anthocyanins; increase in plant
CC anthocyanins, or alteration in light response or shade avoidance. The
CC transgenic plant, polynucleotides and polypeptides are useful in
CC bioinformatic search methods. This is the amino acid sequence of a plant
CC transcription factor, and an orthologue of Arabidopsis thaliana
CC transcription factors isolated in the invention, that can be used in the
CC creation of a transgenic plant with altered traits.
XX
SQ Sequence 969 AA;
XX
Query Match 10.8%; Score 575.5; DB 8; Length 969;
Best Local Similarity 23.2%; Pred. No. 5e-39;
Matches 235; Conservative 161; Mismatches 339; Indels 279; Gaps 40;
Qy 38 KKVLNVNPFKSSKIYREYEVKMTKVLNRRKPGKPPFKKTEIPIIDRAKLFQHL 97
Db 158 KKVMVPAH--HFLVQVADRDLHYDYSINPEVLSK----- 190
Qy 98 RHEKKQTDFILEDYFDEKDTVYVSVCLNTVTTSKMLVSEKVKVKSSE---KKDEKDEKK 154
Db 191 -----TANRNVKMLLVKN---YKDSHLGGKSPAYD-GRK 220
Qy 155 ILTYM--ILTYRKKFHLNFSRENPEKDEANRSYKF-LKNV-----MTQKRYA 200
Db 221 SYLTAGLPFDSEKFEFVNLAERKADSGSGKORPFKAVKXNVVTSTLDYQLQQFLDRKQEA 280
Qy 201 PFVNEIKVQFAKNFVYDNNNSILRVPSFSFHDNRFQSL-----EVAPRIEAWFGI 251
Db 281 PY--DTIQVLDVLRDKPSPNDYVSGRSP-----FHTSLGKDARDGEGELGDIYWRGY 333
```





the polypeptides encoded by the polynucleotide described above. The transgenic plant is useful for producing a plant that has an altered trait e.g. an enhanced tolerance to abiotic stress (increased tolerance to chilling, germination in cold conditions, freezing tolerance, tolerance to heat, tolerance to drought, tolerance to osmotic stress, tolerance to salt, tolerance to phosphate limitation, tolerance to potassium limitation, decreased sensitivity to nitrogen limitation), altered hormone sensitivity, reduced sensitivity to abscisic acid, an altered response to ethylene, disease resistance, altered susceptibility to Botrytis, altered susceptibility to Fusarium, altered susceptibility to Erysiphe, altered susceptibility to Pseudomonas syringae, altered susceptibility to Sclerotinia, altered sugar sensing, improved seed germination and seedling vigor, early flowering, late flowering, extended period of flowering, an inflorescence architectural change, a change in stem bifurcations, a lack of a shoot meristem, reduced meristem cell differentiation, altered phylloclaxy, altered branching pattern, reduced apical dominance, reduced trichome density, ectopic trichome development, altered trichome development, altered stem morphology, increased root growth, increased root hairs, altered seed morphology, increased root proliferation/cell differentiation, premature senescence, delayed senescence, lethality, increased necrosis, an increase in seedling or plant size, decreased plant size, a change in leaf morphology, increased altered leaf development, increased leaf size and mass, glossy leaves, leaf cell expansion, change in seed morphology, altered seed coloration, increased seed size, decreased seed size, altered seed shape, change in leaf biochemistry, increased leaf wax, an alteration in leaf prenly lipid content, increased leaf insoluble sugars, decreased leaf insoluble sugars, increased leaf anthocyanins, an alteration of leaf fatty acid content, an alteration of leaf glucosinolate content, change in seed biochemistry, an increase in seed oil content, decrease in seed oil content, increase in seed fatty acid content, decrease in seed fatty acid content, increase in seed protein content, decrease in seed protein content, alteration in seed prenly lipid content, increase in seed sterols, upregulation of genes involved in secondary metabolism, increase in root anthocyanins, increase in plant anthocyanins, and alteration in light response or shade avoidance. The present sequence represents a thalaeares transcription factor of the invention.

XX Sequence 969 AA;

Query Match 10.8%; Score 575.5; DB 8; Length 969;  
 Best Local Similarity 23.2%; Pred. No. 5e-39;  
 Matches 235; Conservative 161; Mismatches 339; Indels 279; Gaps 40;

QY 38 KVVLLVNFKPSKIDREYVEYVMTKEVLRKPKFPKTEIPIDRAKLFQWHL 97  
 DB 158 KVVVRAN--HFLVQVADRLVHYDVSVINPEVSK----- 190

QY 98 RHEKKQTDILEDYVDEKDTVYVSVCLNTVTTSKMLVSEKVKVCKDSE---KKDEKLEKK 154  
 DB 191 -----TVNRNVNKKLVKN---YKDSHLOGSKSPAYD-GRK 220

QY 155 ILVTM--ILTYRKPHLNFSPENPEKDEANRSYKF-LKNV-----MTQKRVYA 200  
 DB 221 SLYTAGPLPDSKEFVNLAERADSGSGKORPKVAVKXVTTDLVYQOQFLDRKQEA 280

QY 201 PFVNEIKVQFAKVFYDNNLSILRVSPESHDPNRPQSL-----EVAPRIEAWFGI 251  
 DB 281 FY--DTIQVLDVVLDRKPSNDYVSVGRSP-----FHTSLGKADRGKGLGDGYEWRGY 333

QY 252 YIGIKELFDEGEVFNPAVDKLFVNAKPKMSLLDYLLIVDPQSCNDDVRKOLTKLWAGK 311  
 DB 334 FQSLR--LTQNGUSLNDVARSFY-----EPVVVTFISKFLNIRDL--N 375

QY 312 MTRQARPRIRQLLENLKLKCAEVMDNEMSLRHLATFLDLCEENSLVYKVTGKSD-- 369  
 DB 376 RPLRDSRLKVKVLTLLKVLH--WNGTSA-----KISGLSSLP 415

QY 370 ----RGNAAKVDTTLLFKIYEENKPF-IEFPHPLPVKVKSGAKAYAVPMHELVHEKPKR 424  
 DB 416 IRELRFLEDSKSEKTVQVFAEKYKRVKYQALPAIQGTGSDTRPVYLPMLCQIDE-GQR 474

QY 425 YKNRDLVWQDFLKEATRPKDPHYKENTLKMELDFSSSEELNPFVERFGLCSKLQWIECP 484

Db 475 YTRLNEKQVYALLKATCORPPD--RENSIKNLVVKNNYNDLDS--KEFGMSVTTQLASIE 531  
 QY 485 GKVLKEPMLVNSVNEQIKMTPTVIRGFOEKOLNVVPEKELCCAVFVNVNBTAGNPCLENDV 544  
 Db 532 ARVLPPLMKYHDSGKEKVNPRLG-----QWNMDKK----- 564  
 QY 545 VKFYTELIGCKFRGIRIGANENRGAOSIMYDATKNEYAFYKNCNTLNTIGRFEIAATEA 604  
 Db 565 -----MVNGAK---VTSWTCFEPKQPAIPFISCPPEH-----IEEALLD- 600  
 QY 605 KMFPERLPDEQKVLMPILIIISKQLNAYGFVKYCHDTHTIGVANOHITSETVTKALASURH 664  
 Db 601 --IHKRAPGLQ-----LLIVLPDVTGSYKIKRICETELGIVSQCCOPQVNVKL----- 648  
 QY 665 EKSGKRFYQIALKINAKGGINOELDWSBIAEISPEEKERRKTMPL-----DAIRNIPILITRPTIIMGADV 719  
 Db 649 ---NKQYMNVALKINVKTGRTVNL-----DAIRNIPILITRPTIIMGADV 694  
 QY 720 THPTSYSGIDYSIAAVVASIN-FPGTIYRNMIVTQ-----EE-----CRPGERAVAHGR 767  
 Db 695 THPQGEDSSPSIAAVVASMDWPEINKYGLVSAQAHRREIIQDLYKLVQDPQRLVH-- 752  
 QY 768 ERTDILEAKPVKLLRFAENNDRAPAHIVVYRDGVSDEMLRVSHDELAKSEVQKFM 827  
 Db 753 --SGLIREHFIAPRA-----TCQIPQRIIFYRDGVSQGVFSQVLLHEMTAIRKACNSLQ 805  
 QY 828 SERDGDPEPKYFIVIQKHNTLRRLRMKPKVNVNKLTPAETDVAVAVKQWEDMK 887  
 Db 806 -----ENVYPRVTVFVQKHHTLRLPEQHG-----NRDWT----- 836  
 QY 888 ESKETGIVNPSSGTTVDKLVSKYKDFFLASHHGVLTSGRPGHYTVYDDKMGMSODEVY 947  
 Db 837 --DKSGNIQD--GTVVDTKICHNEFDYLNSHAGIQGTSRPAHYHVLLENGPTAQDLQ 892  
 QY 948 KMTYGLAFLSARCKEISLSPVHYAHLSCAKELVYTKHYIGDYAQPRTR 1001  
 Db 893 MLTNNLCTYVARCTKSVSVPPAYAYALAFRA-----RYTMESEMSDGGSSRSR 942

RESULT 11  
 ABP64718  
 ID ABP64718 standard; protein; 859 AA.  
 XX  
 AC ABP64718;  
 XX  
 XX 25-FEB-2003 (first entry)  
 DT  
 XX Human protein SEQ ID 378.  
 DE  
 XX Human; expressed sequence tag; EST; haematopoietic disorder;  
 KW central nervous system disease; viral infection;  
 KW peripheral nervous system disease; non-healing wound; infectious disease;  
 KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;  
 KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;  
 KW antileukemic; antiinflammatory; immunosuppressive; neuroprotective;  
 KW cytostatic; haemostatic; virucide; antibacterial; fungicide;  
 KW immunostimulant; cerebroprotective.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200259260-A2.  
 PN  
 XX 01-AUG-2002.  
 PD  
 XX 16-NOV-2001; 2001WO-US042950.  
 PP  
 XX 17-NOV-2000; 2000US-00714936.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;  
 PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI: 2002-590824/63.  
DR N-PSDB; ABQ99304.  
XX  
PT New isolated polynucleotide, useful in research, diagnostic or  
PT therapeutic methods, e.g. preventing or treating disorders involving  
PT aberrant protein expression or biological activity.  
XX  
PS Claim 20; SEQ ID NO 378; 394pp; English.  
XX  
CC The present invention relates to novel human coding sequences (ABQ99268-  
CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in  
CC therapeutic, diagnostic and research methods. The polynucleotides may be  
CC used in the field of molecular biology as hybridisation probes, primers  
CC for PCR, for chromosome and gene mapping, for the recombinant production  
CC of protein, or in generation of anti-sense DNA or RNA. The  
CC polynucleotides are useful in diagnostics as physical mapping of the  
CC (ESTs) for identifying expressed genes or for physical mapping of the  
CC human genome. The proteins may be used as molecular weight markers, or as  
CC nutritional sources or supplements. The proteins may be used to maintain  
CC and expand cell population in a totipotential or pluripotential state  
CC useful for re-engineering damaged or diseased tissues, transplantation,  
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The  
CC polynucleotides and proteins are useful for preventing, treating or  
CC ameliorating disorders involving aberrant protein expression or  
CC biological activity, e.g. haematopoietic disorders, central/peripheral  
CC nervous system diseases, mechanical and traumatic disorders, non-healing  
CC wounds, immune deficiencies and disorders, infectious diseases caused by  
CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
CC reactions and conditions, coagulation disorders, or cancer. The  
CC polynucleotide sequences of the invention were assembled from ESTs  
CC isolated mainly by sequencing by hybridisation, and in some cases,  
CC sequences obtained from one or more public databases. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 859 AA;

Query Match 10.6%; Score 566; DB 5; Length 859;  
Best Local Similarity 23.1%; Pred. No. 2.7e-38;  
Matches 234; Conservative 160; Mismatches 370; Indels 250; Gaps 39;  
QY 19 PEMKWLARTGKCDGFEYKVKVLLVNWPKFSKIYDREYVEYVQMTKEVLNRKPKPF 78  
DB 17 PIQGYAFKPPRPDPFTSGRTIKLQANFFEMDIPKIDIVHYELDIKPEK-----C 66  
QY 79 PKKTEIPIDRAKLPWQHLRHEKKQTDFILEDY--VFDEKDTVYSVCLNTVTSMKLVSE 136  
DB 67 PRVNVREIVEHM-----VQHFKTQ---IFGDRKPVFDGRKNLYTAMPLPIGRDKV---E 114  
QY 137 KVVVKDSEKKDEKLEKKILYTMILYTKKFKHLNFSRENPEKDEANRSYKFLKNVMTQ- 195  
DB 115 LEVTLPGEGKDR--IPKSIKWVSCVSLQALHDALSGRLSPVFE---TIOALDVVMRHL 169  
QY 196 -KVRVAPFVNEEIKVQFKNFVYDNNLSILRVPSFHDNRFQSLVAPRIEAWFCIYIG 254  
DB 170 PSKRYTPV-----GRSFTASEGCSNP-----LGGREVFWFQFHS 205  
QY 255 IKELFGEPEVLPFAIVDKLFPYNAKMSLDYDILLIVDPOSCNDNDVRKDLTKLMAGKMTI 314  
DB 206 VRPSL-WKMLNLDVSATAFYKA--QPVEFVCEVLDPKSIIEQ-----QKPL 250  
QY 315 QQAARPRIQLENLKLGK---CAEVMDN-----EMSLRTERHLTFLDCENSLVYKVTOK 367  
DB 251 TDSQRVKFTKEIKGLKVEITHCCQMKRKRYVNCVTRRPASHQTF-PLQESGQTVECT-- 307  
QY 368 SDRGRNAKYDITLFIYBENKFIIEPHLPVKKVSGAKAYVPMHELVHEKPKQYKN 427  
DB 308 -----VAQY-----FK--DRHKLVLPVPLPCLQVQEQOKHTYLPLEVCNIVAGRCIYK 355  
QY 428 RIDLVNQDKFLKRAAT-RKPHDYKENTLKMELKELDFSSSEELNFFVERFGLCSKLOMBPCPK 486

DB 356 LTD--NOTSTMIRATARGAPDRQBEISIKLMRSASFNTDP--YVREFGIMVKDEMTDVTGR 411  
QY 487 VLKEPMLVNSVNEQIKMTPV--IRGFQEKQLNV-----VPEKELCAAPVNVNE 532  
DB 412 VLQPPSIIYGRNKAIATPVQGVWDMRNKQFHTGIEIKWAIACAPAPQKQ-CTEVHLKSF 470  
QY 533 T-----AGNPCLEENDVWVKFYTELIGGCKFRGIRIGANENRGAQSI--MYD 576  
DB 471 TEQLRKISRDAGMPIQGPFCF-----CKYA-----QGADSVPEMPFR 506  
QY 577 ATKNEYAFYKNCNTLNTGIRFEIAATEAKNMPERLPDKEQKVLMIIFIIISKRQLNAYGFVK 636  
DB 507 HLKNTYA-----GLQLVVVILPGKTPVYAEVK 533  
QY 637 HYCDHTIGVANOHTSETVTVTKALSLRHEKSGKRIFYQIALAKINAKLGGINGELDWSEIA 696  
DB 534 RVGDTVLGKMATQCVMQKNV-----ORTTPQTLNCLKINVKLGGVNNIL----- 578  
QY 697 EISPEKERRRKTMTVMYGVIDVHTPTVSIGDYSIAAVVASINPGTTIYRNMTVQBEC 756  
DB 579 --LPQGRPPVFOQPV-IFLGADVTHPPAGDGKKPSIAAVVGSMDAHPNRYCATVRVQQ-- 633  
QY 757 RPERAVAHGRERTDILEAKFVKLLREFAENNDRNAPAHIVVYRDCVSDSEMLRVSHDEL 816  
DB 634 -----HROEIIQDLAAMVRELLIQFYKSTRFK-PTRIIFYRDSVSEGOQOVLHHEL 684  
QY 817 RSLKSEVKQFMSERDGEPEPKYTFIVIQKHNTLLRMEKDKPVNWKDLTPAETDVAV 876  
DB 685 LAIREACIKL-----EKDYQPGITTFIVQKRRHTRLF-CTDKNERVKGSGNIPA----- 732  
QY 877 AAVKQWEEEDMKESKETGIVNPSGGTVDKLIYSKYKFDFFFLASHRGVLGTSRPGHYVMY 936  
DB 733 -----GTTVDTKITHPTFEFDYLCSHAGIQGTSRPSHHVLM 769  
QY 937 DDKMGSDQEVYKMTYGLAPLSARCKKPISLPVVHYAHLSCBAKELYTYKEH 990  
DB 770 DDNRFSDELQILTYQLCHTYVYVTRSVISIPAPAYYAHVAPRAR-YHLVDKEH 822  
RESULT 12  
AEF73791  
ID AEF73791 standard; protein; 859 AA.  
XX  
AC AEF73791;  
XX  
DT 06-APR-2006 (first entry)  
XX  
DE Human Argonaute protein hAgo2.  
XX  
KW Argonaute; hAgo2; RNA interference; RNAi; drug screening;  
KW X-ray crystallography; gene silencing; protein structure.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 232..352  
FT /label = PAZ\_domain  
FT Domain 583..832  
FT /label = PIWI\_domain  
FT  
XX WO2006015258-A2.  
XX  
XX PN  
XX  
PD 09-FEB-2006.  
XX  
XX 28-JUL-2005; 2005WO-US027084.  
PF  
XX 28-JUL-2004; 2004US-0592297P.  
PR  
XX 29-JUL-2004; 2004US-0592269P.  
PR  
XX (COLD-) COLD SPRING HARBOR LAB.  
XX  
XX Joshua-Tor L, Song J, Hannon GJ, Liu J, Carmell MA, Rivas P;  
PI Marsden C;

XX WPI; 2006-155768/16.

XX Crystalline Argonaute useful for identifying agent that binds Argonaute

XX protein, utilized for treating pancreatic cancer, treating hepatitis C

PT infection, cancer or inflammatory diseases.

PT Disclosure; SEQ ID NO 3; 215pp; English.

XX

CC The present sequence is that of human Argonaute protein hago2. Argonaute

CC proteins are involved in RNA interference (RNAi). The invention provides

CC a crystalline Argonaute protein, which comprises (i) an N-terminus,

CC middle and PWW domain which form a crescent-shaped base, and (ii) a PAZ

CC domain, which is positioned above the crescent-shaped base, resulting in

CC a cleft between the crescent-shaped base and the PAZ domain. The

CC structure of the full-length Argonaute protein ABF73751 from the

CC archaeobacterium Pyrococcus furiosus (PfiAgO), as determined by x-ray

CC crystallography to 2.25 Angstrom resolution, is provided. The invention

CC also provides an isolated complex comprising an Argonaute protein and a

CC single-stranded RNA hybridized to its target nucleic acid. In certain

CC embodiments, the single-stranded RNA is bound to the PAZ domain of the

CC Argonaute protein, and may further interact with the crescent-shaped base

CC of the Argonaute protein. Also claimed are: a method of determining the

CC three-dimensional structure of an Argonaute protein or its mutant,

CC derivative, variant, analog, homolog, sub-domain or fragment by alignment

CC with the PfiAgO amino acid sequence to match homologous regions; a method

CC of identifying an agent that binds an Argonaute protein by applying a

CC three-dimensional molecular modeling algorithm to the atomic coordinates

CC of an Argonaute protein to determine the spatial coordinates of the

CC binding pocket of the Argonaute protein, and electronically screening the

CC scored spatial coordinates of a set of candidate agents against the

CC spatial coordinates of the Argonaute protein binding pocket; a computer-

CC based method for the analysis of the interaction of a molecular structure

CC with an Argonaute protein; a data array comprising the atomic coordinates

CC of an Argonaute protein; a computer-readable storage medium encoded with

CC the atomic coordinates of an Argonaute protein; an electronic

CC representation of a domain or binding site of the Argonaute protein; a

CC method for obtaining a crystal of an Argonaute protein; a method for

CC identifying an agent that modulates the activity of an RNAi construct by

CC identifying an agent that modulates the expression and/or activity of an

CC Argonaute protein; a method for identifying an agent that potentiates the

CC activity of an RNAi construct by identifying an agent that increases the

CC expression and/or activity of an Argonaute protein; a method of

CC identifying an agent that modulates the activity of an RNAi construct by

CC examining a change in Argonaute protein activity in the presence of a

CC candidate agent; a composition for targeted gene inhibition comprising an

CC agent that modulates the RNAse activity of an Argonaute protein; a cell

CC line that overexpresses an Argonaute protein; an assay for identifying

CC nucleic acid sequences for conferring a particular phenotype in a cell; a

CC nucleic acid composition composed of a nucleic acid comprising an RNAi

CC construct and a nucleic acid encoding an Argonaute protein, where the

CC RNAi construct comprises a nucleic sequence encoding a single-strand

CC short interfering RNA (siRNA); a pharmaceutical composition comprising

CC the nucleic acid composition; and a cell expressing the nucleic acid

CC composition. The methods and compositions of the invention are useful for

CC enhancing the effectiveness of an RNAi therapeutic.

XX Sequence 859 AA;

SQ

Query Match 10.6%; Score 566; DB 10; Length 859;

Best Local Similarity 23.1%; Pred. No. 2.7e-38;

Matches 234; Conservative 160; Mismatches 370; Indels 250; Gaps 39;

QY 19 PEMKWLARPTGKCDGKFKYKVVLLVNWFKSSKIYDREYVEYVMTKEVLNRKPKGP 78

DB 17 PIQGYAFKPPRPDPFGTSGRITKIQANFPFEMDIPKIDIVHYELDIKPEK-----C 66

QY 79 PKTETPIPDRAKLFWQHLNHEKKQTDFILEDY--VFDEKQTVYSVCLRNVTWVKMLVSR 136

DB 67 PRVNRNREIHEM-----VQHFKTQ---IFGDKRPVDFGKRNLYTAMPLPIGRDKV--E 114

QY 137 KVVYKQDEKDKLKKILYTMILYRKKFHLNFSRENPEKDEANRSVKFLKNVMTQ- 195

DB 115 LEVTLPEGKDR--IFKVISIKWVSCVSLQALHDALSGRLPSVPFE-----TIOALDVNRHL 169

QY 196 -KRYAPFVNEEIKVQPAKNFVYDNNISILRVPSFSDHPNRFPEQSLEVAPRIEAMFGIYIG 254

DB 170 PSMRYTPV-----GRSFFTASGCSNP-----LGGGREVWFGPHQS 205

QY 255 IKELFDGEPVNFVVDKLFYNAPKMSLLLDYLLIVDPQSCNDDVDRKDLKTKLMAGKWTI 314

DB 206 VRPSL-WFOMMLNIDVSATFYKA--QPVIEFVCEVLDPKSIEEQ-----OKPL 250

QY 315 ROAARPRIOLLENLKLK---CABVDN---BMSRLTERHLTFLDCEENSLVYKVTGK 367

DB 251 TDSQKVFTKEIKGLKVEITHCGOMKKYKVCNVRPASHQTP-PLQSGSQVTECT-- 307

QY 368 SDRGRNAKDYDTTLFKIYEENKKFIEPPLPLVKVKSAGKEYAVPMHEHLEVHEKPPQRYKN 427

DB 308 -----VAQY-----FK--DRHLVLRYPHLPCLQVGOEQKHTYLPLEVCNVAGORCIKK 355

QY 428 RIDLVMDQKFLKAT-RKPHDYKENTLKMUKELDFSSSEELNFVERFGLCSKLQMIKCPGK 486

DB 356 LTD--NQTSMIRATARSAPDROEISKLMSASFNTDP--YVREFGIMVKDEMTDVTGR 411

QY 487 VLKEPLMNVSNVSOIKMTVP--IRGFOEKOLNV-----VPEKELCCAFVYVNE 532

DB 412 VLOPPSILYGRNKAIATPVQGVMDNRNKQFHTGIEIKVWAIACFAPQRO-CTEVHLKSF 470

QY 533 T-----AGNPLEENDVVVKFYTELIGCKPRGIRITGAMENRGAOSI--MYD 576

DB 471 TEQLRKISRDAGMPIQGPCF-----CKYA-----QGADSVEMFMR 506

QY 577 ATKQVAFYKNCYTLNTGIRGFEIATAEKNMFERLPDKQKVLMPFIISIKROLNAYGFVK 636

DB 507 HLKNTYA-----GLQLVVVILPKTTPVYAEVK 533

QY 637 HYCDHTIGVANOHTSETVTTKALASLRHKGSKRIFFQIALKINAKLGGINQELDWSEIA 696

DB 534 RVGDTVLGWTATQCVQKNV-----QRTPTQSLNCLKINVLGGVNNIL----- 578

QY 697 EISPEEKERRKMTPLTMYGIDVTHTPTSYSGIDYSAAVVASINPGGTIYRNMIVTQEEC 756

DB 579 --LPQGRPPVFPQPV-IFLGADVTHPPAGDGKKPSIAAVVGSMDAHPNRYCATRVQQ-- 633

QY 757 RGERAVAIGRERTDLEAKFVKLLREPAENNDNRAPAHIVVTRDGVSDSEMLRVSHDEL 816

DB 634 -----HROEIIQDLAAVRELLIQFYKSTRFK-PTRIIFYRDGVSEGGQQOVLHHEL 684

QY 817 RLSKSVKQFMSERDEGPEPKYTFIVIOKRHTLLRMEKDKPVVNKDLTAEITDVAV 876

DB 685 LAIREACIKL-----EKDQPGITFIVOKRRHRLFP-CTDKNRVKGSGNIIPA----- 732

QY 877 AAVKQWEEDMKESKETGIYNPSSGTTVDKLVISKYKFPDFFLASHHGVLGTSRFGHTYVMY 936

DB 733 -----GTTVDTKITHPTDFDFYLCSHAGIQGTSRFSHYHVLW 769

QY 937 DKGMSQDEYKMTYGLAFLSARCCKPISLPVPVHYAHLSCSEKAKELYRTYKXH 990

DB 770 DDNRFSSEDLQILTYQLCHTYVVRCSVSIPAPAYAHLVAFRAR-YHLVDKEH 822

RESULT 13

ABG97473

ID ABG97473 standard; protein; 860 AA.

XX ABG97473;

XX AC ABG97473;

DT 16-DEC-2002 (first entry)

XX Human NAAP7 from INCYTE no.1725129CD1.

XX DE

XX Human; nucleic acid associated protein; NAAP; cancer;

KW cell proliferative disease; cancer; atherosclerosis; hepatitis;

KW neurological disorder; Parkinson's disease; Alzheimer's disease; stroke;

KW epilepsy; developmental disorder; renal tubular acidosis; anaemia;







PF 23-MAR-2003; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
PA (PEKE ) PE CORP NY.  
XX  
XX  
PI Venter JC, Adams M, Li PWD, Myers BW;  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL11149.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 27930; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 950 AA;

Query Match 10.5%; Score 560.5; DB 4; Length 950;  
Best Local Similarity 22.9%; Pred. No. 9.2e-38;  
Matches 225; Conservative 170; Mismatches 386; Indels 201; Gaps 35;  
  
QY 38 KVVLLVNNFKSSKIYDREY-YEYEVKMTKVLNRKPKPKPKTEIPIDRAKLFQW 96  
DB 112 RPILVRANHFQVT---MPRGYVHHYDINIQDKCPKRVNREI---IETVHAYSKIFG-- 163  
  
QY 97 LRHEKQTDFFILEDYVDFDEKDTVYSVCRNLNTVTSKMLVSEKVKVKKDSEKKDKLEKKIL 156  
DB 164 -----VLKP-VFDGRNNLYTRDPLPIGNERL---ELEVTLPGEGKDR----- 201  
  
QY 157 YTMILTYRKFFHLNFSRENPEKDEEANSYKFLKNVMTOKYAPVPVNEEIKVQAKNFV 216  
DB 202 ---IFRVTIKWAQVSLFLEALEGR-----TRQIPYDAILALDVVMVRLPSMT 248  
  
QY 217 YD--NNSILRVPSFHPNRFQSLVAPRIEAWFGIYIGIKELPDGEPVNFPAIVDKLF 274  
DB 249 YTPVGRSFPSSPEGYHP-----LGGREWFVGFHQSVRP-SQWKMLNIDVSATAF 299  
  
QY 275 YNAPKMSLLDYLLIYDPOSNDVVRKDLTKLMAGKMTIRQAARPIRQLLENKCLKCA 334  
DB 300 YKA--QPVDFMCEVLDIRDINEQ-RKPL-----TDSQRVKFTKEIKGLKIEIT 345  
  
QY 335 EVDNEMSLRTERHLTFLDCEENSLVYKVTGSDRG-----RNAKYDITLFLKI 384  
DB 346 H-----CGQMRKRYVCNVTRPAQMQSPFLQLENGQTVECTVAKY 386  
  
QY 385 Y-EENKKFTEFPHLPVVKSGAKEVAVPMHELVHEVKPQRYKNRIDLVMDQKFLKRA 443  
DB 387 FLDKVRMKLRYPHLPCLQVQGEHKHTYLFLEVNI-VAGQRCIKKLTDMQSTMIKATAR 445  
  
QY 444 KPHDYKENTLKMKELDFFSBEELNFERFGLCSKLQMECPGKVLKEPMLVNSVNEIQIK 503  
DB 446 SAPDREREINLVKRADFNND--SYVQEFGLTISNMSMEVRGVLPPPKLYQYGR----- 498  
  
QY 504 TPVIRGFQEKQLNVVPEKELCCAVFVNVTAGNPNCLEENDVVKFYTELIGGCKFRGIRG 563  
DB 499 --VSTGLTGQQL-FPPQNKVSLA-----SPNQGVMDMRGKQFFTG-----EIRIW 541  
  
QY 564 ANENRGAQ-SIMYDATKNEYAPYKNTCTLNTG---IG-----RFEIATAEAKMFERLPDK 614

Search completed: July 5, 2006, 13:28:31  
Job time : 206 secs

DB 542 AIACFAPQRTVREDALRNFTQLOKISNDAGMPIIQPCFCCKYATGPDQVPMFRYLKIT 601  
QY 615 EOKVLMFIIISKRLNAYGFVKHYCDHTIGVANQHITSETVTYKALASLRHEKSGKRIFYQ 674  
DB 602 FPGQLQVAVVLPKTPVYAEVKRVGDTVLGMATQCVQAKNVNKT-----SPOTLSN 652  
QY 675 TALKINAKLGGINQELDMSAEISPEEKERRKTPMLTMVYGDIDVTHPTSYSGIDYSIAA 734  
DB 653 LCLKINVLGGINSIL---VPSIRPKVFNBP-----VIFLGADVTHPPGDNKKPSIAA 703  
QY 735 VVASINPGGTIYRNMIVTQECCRPGERAVAHGRERTDILEAKFVKLLREFAEENNDRAPA 794  
DB 704 VVGSDAHPRSRYAATVRVQQ-----HQEIIQELSSMVRELLIMFYKSTGCGYKPH 753  
QY 795 HIVYRDGVSDSEMLRVSHDELRLSLKSEVKQPMSESDGEDPE--PKYTFIVIOKRNHTRL 852  
DB 754 RIILYRDGVSEGFPHVLQHELTAREACIKL-----EPEYRPGITFIVVQKRHHTRL 806  
QY 853 LRRMEKDKPVVVKDLTPAETDVAAVAAQWEEDEKSKETGIVNPSGGTTVDKLIIVSKYK 912  
DB 807 FCAEKKK-----QSGKSG--NIPAGTIVDVGITHPT 836  
QY 913 FDFFLASHHGVLGTSRPGHYTVMYDDKGMSEDEVYKMTYGLAFLSARCKRPISLPVPHY 972  
DB 837 FDFYLCSHQGIQTSRPSHYHVLWDDNHFDSDLOCLTVQLCHTVYVRCRSVSIPAPAYV 896  
QY 973 AHLSCAKELYRTTYKEHYIGD 994  
DB 897 AHLVAFRAR-VHLVEKEHDSGE 917



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 13:28:51 ; Search time 48 Seconds  
(without alignments)  
2044.606 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 5349  
Sequence: 1 MSNPFPELEKGFYRHSIDPE.....RHEMEHFLQTNVYKPGMSFA 1020  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:.\*  
1: Pirl:.\*  
2: Pirl2:.\*  
3: Pirl3:.\*  
4: Pirl4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4677.5	87.4	905	2 T23510	hypothetical prote
2	604.5	11.3	997	2 A84578	Argonaute (AGO1)-1
3	557	10.4	1000	2 T22391	hypothetical prote
4	556	10.4	813	2 JC6569	translation initia
5	546.5	10.2	1032	2 T23164	hypothetical prote
6	546.5	10.2	1035	2 T23165	hypothetical prote
7	540.5	10.1	891	2 T32079	hypothetical prote
8	537	10.0	958	2 S41013	hypothetical prote
9	537	10.0	1040	2 D88568	protein ZK757.3 li
10	510.5	9.5	1014	2 H86438	protein T19823.7 li
11	488	9.1	988	2 T52134	Zwille protein [im
12	484	9.0	834	2 T41568	argonaute-like pro
13	473	8.8	898	2 T25164	hypothetical prote
14	428.5	8.0	930	2 A84668	Argonaute (AGO1)-1
15	402	7.5	887	2 T01113	translation initia
16	356	6.7	185	2 T27784	hypothetical prote
17	347.5	6.5	1139	2 T33275	hypothetical prote
18	311.5	5.8	674	2 T19268	hypothetical prote
19	311.5	5.8	1030	2 T16114	hypothetical prote
20	309.5	5.8	722	2 T30995	hypothetical prote
21	302.5	5.7	990	2 F88925	protein T22H9.3 li
22	296.5	5.5	965	2 T22933	hypothetical prote
23	294	5.5	1032	2 T19324	hypothetical prote
24	287.5	5.4	824	2 T20351	hypothetical prote
25	272	5.1	1057	2 T16676	hypothetical prote
26	259.5	4.9	939	2 T18974	hypothetical prote
27	247.5	4.6	881	2 T31818	hypothetical prote
28	237	4.4	945	2 T23965	hypothetical prote
29	223	4.2	892	2 T15195	hypothetical prote

30	211.5	4.0	944	2 T18911	hypothetical prote
31	206.5	3.9	697	2 T15179	hypothetical prote
32	192.5	3.6	2269	2 T28677	rhopty protein -
33	190.5	3.6	946	2 S28061	SCP1 protein - rat
34	183	3.4	363	2 T34486	hypothetical prote
35	182.5	3.4	587	2 T34339	hypothetical prote
36	175.5	3.3	1979	2 C71622	hypothetical prote
37	175	3.3	2253	2 T30336	nuclear/mitotic ap
38	175	3.3	2401	2 T28676	rhopty protein -
39	174.5	3.3	1302	1 JC6009	surface-located me
40	172.5	3.2	1650	2 A84647	hypothetical prote
41	172	3.2	1939	2 T18372	repeat organelle
42	170.5	3.2	324	2 G87774	protein C24A11.3 l
43	170	3.2	1170	2 A72287	hypothetical prote
44	169.5	3.2	2829	2 A42771	reticulocyte-bind
45	168.5	3.2	980	2 E71606	hypothetical prote

ALIGNMENTS

RESULT 1

T23510  
hypothetical protein K08H10.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T23510  
R;Gardner, A.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z19750  
A;Accession: T23510  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-905 <WIL>  
A;Cross-references: UNIPROT:Q9XU82; UNIPARC:UPI0000017BAA9; EMBL:Z83113; PIDN:CAB05546.1  
A;Experimental source: clone K08H10  
C;Genetics:  
A;Gene: CBSP:K08H10.7  
A;Map position: 5  
A;Introns: 19/3; 86/2; 190/3; 209/2; 269/1; 341/2; 468/3; 671/3; 832/3

Query Match	87.4%;	Score	4677.5;	DB	2;	Length	905;
Best Local Similarity	88.7%;	Pred. No.	1e-254;				
Matches	905;	Conservative	0;	Mismatches	0;	Indels	115; Gaps 1;
QY	1	MSNPFPELEKGFYRHSIDPEMKWLARPTGKCDKGFYKVKVLLVNNWPKFSSKIYDRYYE	60				
DB	1	MSNPFPELEKGFYRHSIDPEMKWLARPTGKCDKGFYKVKVLLVNNWPKFSSKIYDRYYE	60				
QY	61	YEVKMTKEVLNRKPGKPPPKTETIPIDRAKLFQWHLRHEKKQTDPILEDYVFEKQDTVY	120				
DB	61	YEVKMTKEVLNRKPGKPPPKTETIPIDRAKLFQWHLRHEKKQTDPILEDYVFEKQDTVY	120				
QY	121	SVCLRLTVTSKMLVSEKVKVKKSEKDEKLEKKILYTMILTYRKGFHLNFSRENPEKDE	180				
DB	121	SVCLRLTVTSKMLVSEKVKVKKSEKDEKLEKKILYTMILTYRKGFHLNFSRENPEKDE	180				
QY	181	EANRSYKFLKNVMTQKRVAPFVNEEIKVQFANFVYDNNISILRVPSFHPDRPFQOSLE	240				
DB	181	EANRSYKFLKNVMTQKRVAPFVNEEIKVQFANFVYDNNISILRVPSFHPDRPFQOSLE	240				
QY	241	VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLIYVDPQSCNDDVR	300				
DB	241	VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLIYVDPQSCNDDVR	300				
QY	301	KDLTKLMAGKMTIQAARPRITROLLENKLKCAEYVDNEMSLRTERHLTFLDLCENSL	360				
DB	301	KDLTKLMAGKMTIQAARPRITROLLENKLKCAEYVDNEMSLRTERHLTFLDLCENSL	360				
QY	361	VYKVTGKSDRGNAKKYDITTLFKIYBENKFFIEFPHLPLVKVKSAGKAYAVPMHELVHE	420				
DB	361	VYKVTGKSDRGNAKKYDITTLFKIYBENKFFIEFPHLPLVKVKSAGKAYAVPMHELVHE	420				

```
QY 421 KPORYKNRIDLVQDKFLKRAATKPHDYKENTLKMELKELDFSSSEELNFVERFLGCSKLM 480
DB |||||
QY 421 KPORYKNRIDLVQDKFLKRAATKPHDYKENTLKMELKELDFSSSEELNF 468
DB |||||
QY 481 IBCPGKVLKEPMLVNSNQIKMTVPVIRGQEBKQLNVVPEKELCCAVFVNNETAGNPCLE 540
DB |||||
QY 541 ENDDVVKFYTELIGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIRPEFA 600
DB |||||
QY 601 ATEAKNMPERLPDKEQKLMFIIISKRLNAYGFVKHYCDHTIGVANQHITSETVTKALA 660
DB |||||
QY 486 ATEAKNMPERLPDKEQKLMFIIISKRLNAYGFVKHYCDHTIGVANQHITSETVTKALA 545
QY 661 SLRHEKSGRIFYQIALKINAKLGGINQELDWSEIAEISPEEKERRKTMPLTMYGIDVT 720
DB |||||
QY 546 SLRHEKSGRIFYQIALKINAKLGGINQELDWSEIAEISPEEKERRKTMPLTMYGIDVT 605
QY 721 HPTSGIDYSIAAUVASINPGGTYIRNMIVTQEBRCRGERAVAGHRETDILEAKFKVL 780
DB |||||
QY 606 HPTSGIDYSIAAUVASINPGGTYIRNMIVTQEBRCRGERAVAGHRETDILEAKFKVL 665
QY 781 LREFANNDRAPAHIVVYRDGVSDSEMLRVSHDELRLSLKSVKQFMSEKEDGEDPEPKYT 840
DB |||||
QY 666 LREFANNDRAPAHIVVYRDGVSDSEMLRVSHDELRLSLKSVKQFMSEKEDGEDPEPKYT 725
QY 841 FVIOQRHNTLLRRMEKDKPVVKNKDLTPAETDVAVAQKQWEEDMKSKSETGIVNPSGG 900
DB |||||
QY 726 FVIOQRHNTLLRRMEKDKPVVKNKDLTPAETDVAVAQKQWEEDMKSKSETGIVNPSGG 785
QY 901 TVVDKLVSKYKDFPFLASHHGVLGTSRPGHYTMVYDDKGMQSDDEVYKMTYGLAFLSARC 960
DB |||||
QY 786 TVVDKLVSKYKDFPFLASHHGVLGTSRPGHYTMVYDDKGMQSDDEVYKMTYGLAFLSARC 845
QY 961 RKPISLPVPVHYAHLSCAKELRYKHEHYIGDVAQPRTRHEMEHFLQTNVYKYPQMGA 1020
DB |||||
QY 846 RKPISLPVPVHYAHLSCAKELRYKHEHYIGDVAQPRTRHEMEHFLQTNVYKYPQMGA 905

RESULT 2
A:Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: A84678
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84678
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-997 <STO>
A:Cross-references: UNIPROT:Q9SJK3; UNIPARC:UPI00001256F1; GB:AE002093; NID:94510428; PI
C:Genetics:
A:Gene: At2G27880
A:Map position: 2
C:Superfamily: rabbit translation initiation factor eIF-2C
```

```
Query Match 11.3%; Score 604.5; DB 2; Length 997;
Best Local Similarity 23.8%; Pred. No. 4.5e-26;
Matches 242; Conservative 164; Mismatches 353; Indels 259; Gaps 41;
QY 38 KVVLLVNVKFPSSKIYDREYVEYKMTKEVLNRPKPGKPPPKTEIIPDRAKLFWQHL 97
DB |||||
QY 98 RHEKQTDPILEDYVDEKOTVYVCRNLNTVTSKMLVSEKVVKQDSE---KKDEKLEBK 154
DB |||||
```

```
DB 191 -----TVNRNVKMLLVKN-----YKSHLGKSKSPAYD-GRK 220
QY 155 ILYTM--ILYTRKKFHLPNSRENPEKDEANRSYKF-LKV-----MTQKRYA 200
DB |||||
QY 221 SLYTAGLPFPDSKEFVNLAEKRADGSGSKDRPFKVAVKNVSTDLQYQQLDRKOREA 280
QY 201 PPNVEIEIKVQAKNFVYDNNISILRVPESPHDPNRRFEQSL-----EVAPRIEAWFGI 251
DB |||||
QY 281 PY-DTIQVLDVVLBDKFSNDYVSGRSF-----FHTSLGKADRGREGELGDIWIRGY 333
QY 252 YIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLVDPQSCNDDVRKDKLTKLMAGK 311
DB |||||
QY 334 FQSLR-LAQMLGLSLIDVSARSFY-----EPIVWTDPIFKELNIRDL--N 375
QY 312 MTIROAARPRITQLLENLKLKCAEYVDNEMSLRTERHLTFLDLCRENSIVYKVTGSKD-- 369
DB |||||
QY 376 RPLRSDRLKVKVKVRLTKVLLH-WN-----CTKSA---KISGISSLP 415
QY 370 ---RGRNAKYDITLTKIYEENKKF-IEFPHLPLVKVKSGAKEYAVPMHEHLEVEKPOR 424
DB |||||
QY 416 IRELFTLEDKSEKTVQVFAEKYNYRVKYQALPAIQGSDTRPVVLPMLCOIDE-QOR 474
QY 425 YKNRIDLVQDKFLKRAATKPHDYKENTLKMELKELDFSSSEELNFVERFLGCSKLMQIECP 484
DB |||||
QY 475 YTKRLNEKQVTTALLKATCORPPD-RENSIGNLVKNYNDLDS---KEFGMSVTTQLASIE 531
QY 485 GKVLKEPMLVNSNQIKMTVPVIRG-----FQEKQLNVVPEKELCCAVFVNNETAGNPCLE 540
DB |||||
QY 532 ARVLPPMLKYHDSGEKQWNPRLGQWNNIDKQWNGAKVTSWTCVSFSTRIDRGLP--- 588
QY 541 ENDDVVKFYTELIGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIRPEFA 600
DB |||||
QY 589 ---QEFCKQLIGMVCVKGM-----EFKQPAIPFISCPPEH-----IEEA 625
QY 601 ATEAKNMPERLPDKEQKLMFIIISKRLNAYGFVKHYCDHTIGVANQHITSETVTKALA 660
DB |||||
QY 636 LLD---IHKRAPGLQ---LLVILPDDVTGYSYKIKRICETELGIVSQCCQPRQVKNL-- 676
QY 661 SLRHEKSGRIFYQIALKINAKLGGINQELDWSEIAEISPEEKERRKTMPL-----TMVY 715
DB |||||
QY 677 -----NKQYMNVALKINVKTGGRNTVLN-----DAIRRNPLITDRPTIIM 718
QY 716 GIDVTHPTSGIDYSIAAUVASIN-PGGTYIRNMIVTQ---EE-----CRPGERAV 763
DB |||||
QY 719 GADVTHPQPGEDSSPSIAAUVASMDWPEINKYRGVLVSAQAHRREEIIQDLYKLVQDPQRL 778
QY 764 AHGRETDILEAKFVKLLREFANNDRAPAHIVVYRDGVSDSEMLRVSHDELRLSLKSEV 823
DB |||||
QY 779 VH-----SGLIREHFIAPRA-----TGQIPORIIFYRDGVSGQFSQVLLHEMTAIRKAC 829
QY 824 KQPMSEKEDGEDPEPKYTIVIOQRHNTLLRRMEKDKPVVKNKDLTPAETDVAVAQKQWE 883
DB |||||
QY 830 NSLQ-----ENYVPRVTFVIOQRHNTLLRRMEKDKPVVKNKDLTPAETDVAVAQKQWE 864
QY 884 EDMKESKETGIVNPSGGTVDKLVSKYKDFPFLASHHGVLGTSRPGHYTMVYDDKGMQSD 943
DB |||||
QY 865 -----DKSGNIQIP-GTVVDTKICHPNFEFDYLNSHAGIQGTSRPAHVHVLDDENGFTA 916
QY 944 DEVYKMTYGLAFLSARCCKPISLPVPVHYAHLSCAKELRYKHEHYIGDVAQPRTR 1001
DB |||||
QY 917 DQLQMLTNLCYTVARCTKSVIYPPAYYLAAPRA-----RYNMESEMSDGGSSRSR 970

RESULT 3
T22391
hypotheical protein F487.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22391; T23234
R:Coles, L.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19559
A:Accession: T22391
```

[illegible]

```

Db      910 VLWDDNNLTADELQQLTYQMCHTYVCRTRSVISIPAPAYIAHLVAPRAR-YHLVDREHDSG 968
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Qy      994 DYAOQ 998
      :::: ::::
Db      969 EGSQP 973
      :::: ::::

RESULT 4
JC6569
translation initiation factor eIF-2C - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 09-Jul-2004
C:Accession: PC6505; JC5569
R:Zou, C.; Zhang, Z.; Wu, S.; Osterman, J.C.
Gene 211, 187-194, 1998
A:Title: Molecular cloning and characterization of a rabbit eIF2C protein.
A:Reference number: JC6569; MUID:98267198; PMID:9602122
A:Accession: PC6505
A:Molecule type: protein
A:Residues: 336-359;380-409;694-711 <ZOU>
A:Cross-references: UNIPROT:O77503; UNIPARC:UPI00001790FE; UNIPARC:UPI000017900F
A:Experimental source: liver
A:Accession: JC6569
A:Molecule type: mRNA
A:Residues: 1-813 <ZOU2>
A:Cross-references: UNIPARC:UPI0000120D7E; GB:AF005355; NID:g3253158; PIDN:AAC3
C:Superfamily: rabbit translation initiation factor eif-2C
C:Keywords: liver

Query Match      10.4%; Score 556; DB 2; Length 813;
Best Local Similarity 23.6%; Pred. No. 1.8e-23;
Matches 230; Conservative 152; Mismatches 346; Indels 248; Gaps 40

Qy      57 EYTEYEVMTKEVLNRIPGPGPPKKEIPIPDRAKLFQWHLRHEKQTFDILEDY-VFD 114
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      7 DIHYE-----LDIKPEK-CPRRVNRIVEHM-----VQHFKAQ---IFGDRKPVFD 49
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      115 EKDTIVSVCRILNTVTSKWLVSKEVKKVKKDSKKDEKLEKKILYTMILTYRKKFHLNFSRE 174
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      50 GRKNLYTAMPPLTGREKV-----ELEVTLPGGKOR--IPKVSIRKVVSCVSLQALHDLSGR 104
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      175 NPEKDEEANSRYKFLKNVMTQ--KVRYPAPFNVEEIKVQFAKNFVYDNNLSILRVPESHDP 232
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      105 LPSVPFE--TIQALDVVMEHLSMRYPV-----GRSPPTASEGCSNP 145
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      233 NRPEQSLEVAPRIEAMPGIYIGIKELPDGEPVLNFAVDKLFYNAPKMSLDYLLLIQVP 292
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      146 -----LGGGREVVEFGHQSVRPSL-WQWMLNIDVSATAFYKA--QPVIEFVCEVLDF 194
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      293 QSCNDDVRKDKTKLMAGKMTIRQAPPRIRQLLENLKLK---CAEVDN-----EMSLRT 345
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      195 KSIEQ-----QKPLTDSQVKETKEIKGLKVEITHCGQMKRYKRVNCNVTRRP 242
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      346 ERHLTFLDLCEENSLVYKVGKSDRGNAKKYDTTLFKIYEENKKEFTFPHLPVLVKVKG 405
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      243 ASHQTP-PLQESQTVECT-----VAQY---PK--DRHKLVLRYPHLPCLQVQE 287
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      406 AKYAVPMMEHLEVHEKPPQRYKNRIDLVMDQKFLKGRAT-RKPHDYKENTLQWLKELDFSS 464
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      288 QKHTYLPLEVCNIVAGORCIKCLTD--NOTSTMTIRATARSAPDQEEISIKLMRSASFNTD 345
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      465 ELNFVEREGCLSKLQMECEPGKVLKSPMLVNSVNEQIKMTPV--IRGFQEKQLNV----- 517
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      346 P--YVREFGLMVKDEMTDVTGRVLQPPSILYGGRNKAIAATPVQGVDMRNKQFHTGIEIK 403
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      518 -----VPEKELCCAFVFNVE-----AGNPCLSEENDVVKFYTELIGCK 556
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      404 VWAIACFAPQRQ-CTEVLKSFTEQLRKISRDAGMPIQGQPCP-----CK 447
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Qy      557 FRGIRIGANENRGAQSI--MYDATKNEYAFYKNTLTNTGTGIRFEIATEAKNMFERLPDK 614
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      448 YA-----OCAGDSVGMFRHLKNTYA----- 467
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

```





Db	512	GMPIIGTPEKCYASGVEQVE-----PMFKYLKQTSYSAIQIIVVLPKGTPIYAEVKR	564
Qy	638	YCDHTIGVANOHTSETVTYKALASLRHEKSGKRIFYQIAIKINAKLGGINQBLDSEIAE	697
Db	565	VGDTVLGIATQCVOAKNAIRT-----TPQTLNLCIKMNVKLGVSIL-----	608
Qy	698	ISPEEKEREKTMPLTMYGIDVTHPTSYSGIDYSYAAVVASINPGGTIYRNIMVTOBECR	757
Db	609	-LPNVRPRTFNEBFV-IFLGCDDTHPAAGDTRKPSIAAUVGSDAHPSPAAATVRVQQ----	663
Qy	758	PGERAVAHGRERTDILEAKFVKLLREFAEFNNDNRPAHIVVYRDGVSDSEMLRVSHDEL	817
Db	664	-----HQEIIITDLYMVRELLVQFYENTREF-PARIVVYRDGVSEGLFNVLQYELR	715
Qy	818	SLKSEVKQPMSEBDBPDKYTFIVIQKHNTLLRRMEKDKPVVKNKDLTPAETDVAVA	877
Db	716	AIREACVMLES-----GYQGGITFAVQKRHHHTLFAADKADQ--VGK-----	756
Qy	878	AVKQWEEDMKESKETGIVNPSSGTTVDKLIIVSKYKDFFLASHHGVLGTSRPHYTMVD	937
Db	757	-----AFNIPPGTTVDVGITHPTPEDFFLCSHAGIQTSPSRSHVHLWD	800
Qy	938	DXGMSODEVYKMTYGLAFLSARCKPISLPVPVHVHAHLSCEKAKELYRTYKSHYIGDY-A	996
Db	801	DNDLTADQLTYQWCHTYVCTRVSISIPAPAYIAHLVAFRAR-YHLVDRDHGSGEGS	859
Qy	997	QP 998	
Db	860	QP 861	
RESULT 8			
S41013			
hypothetical protein ZK757.3 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 31-Dec-2004			
C:Accession: S41013			
R:Thomas, K.			
submitted to the EMBL Data Library, December 1993			
A:Reference number: S41011			
A:Accession: S41013			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-958 <THO>			
A:Cross-references: UNIPROT:Q9TW94; UNIPARC:UPI0000179102; EMBL:Z29121			
C:Genetics:			
A:Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1			
Query Match			
Best Local Similarity 10.0%; Score 537; DB 2; Length 958;			
Matches 222; Conservative 170; Mismatches 375; Indels 262; Gaps 39;			
Qy	45	NMFKFSKTIYDREYIEYEMTKVNLNRKPGFPFKTEIPIDRAKLFWQ-----HLR 98	
Db	22	NFFAMDKNPKMVIQYHVEI-----HHPGCKLKDQEMRI-----IFWKAVSDDHPNIF 70	
Qy	99	HEKKQTDFILEDYVDFEKTQVTVSVCRL---NTVTSKMLVSEKVKVKDKSEKDEKLEKXI 155	
Db	71	HNK-----FAL---AYDGAHQLYTVARLEFPDDQGSVRLDCEASLPKDNDRTRCAISIQN 123	
Qy	156	LYTMILTYRKFFHLNFRSRENPEKDEANRSYKFLKNMTQKRYAPFVNNEIKVQFAKNF 215	
Db	124	VGPVLEMQR-----TRTN-NLDERVLTPIQLDIIICQSIT-CPLLQN-----SANF 169	
Qy	216	VYDNNILRVPSFHDPNRFEQSLEAVPRIEAWFIYIGIKELFDGEPVFNFAIVDKLPY 275	
Db	170	YTMKSSCYRIPTA-----AGQALDEGGKEMWTFGSSAHIASNRYRPLNDVAHTAFY 223	
Qy	276	NAPKMSLLDYLIIVD-----	293
Db	224	KT-RITVLOFMCVDVLNERTSKPNRNNPRPGPGGPGGYRGGRGGRGGSGYNGFNGRGP 282	
Qy	294	SCN-----DVRKD-----LTKTLMAGK---MTIRQAARPRIRQLLENLKL 331	

Db	283	GANVRDDFGNGLTTMTDLSRDTQLSSFETRI FGDAIRGMKIRAAHRP-----	331
Qy	332	KCAEYMDNEMSRSLTERHLTFLDLCEBNSLVYKVTGKSDRGRNAKKYDTTLTKIYEENKKF	391
Db	332	NAIRVVYKNSLQLPADKLMFQGI DEGRQV--VCSVAD-----YFSEKYGP 375	
Qy	392	IEFPHLPLVKKVSGAKAYAVPMEHLEVHEKPORYKMRIDLVMQDKFLKQATRKPHDYKEN	451
Db	376	LKYPKLPCPLHVGPPTRNTFLPMEHCLI-DSPQYKNKMSEKQTSATIIKAAAVDATOREDR	434
Qy	452	TUKMLKELDFSEELNFVERFGLCSKLOMIECPGKVLKEPMLVNSNEQIKMTPVIRGFG	511
Db	435	IKQLAAQASFGTDP--FLKEFGVAVSSQMIQTARVIQPP-----PIFNGN 479	
Qy	512	EKQLNVVPEKELCCAVF-----VVNETAGNP--CLEENDVY---KPYTELIGCK--	556
Db	480	NRSVNPV-----VFPKDGSWTMDNQTLYPATCRSYSMIALVDPDRQTSLQTFQCSL	531
Qy	557	-PRGIRIGANENRGAQSIMYDATKNEYAFYKNCNTLNTGIG-RFEIAATEAKNMFERLPDK	614
Db	532	TMKATAMGNFPFRWPDLVKYGRSKEDV-----CTLFTEIADBYRVTNTVCD-----	577
Qy	615	EQKVLMTFIIISKRQLNAYGVFVKHYCDHTIGVANQHTSETVTKALASLRHEKSGKRIFYQ	674
Db	578	-----CIIIVLQSKNSDIYMTVKEQSDIVHGINSQCVLKMNVRP-----TPATCAN	624
Qy	675	TALKINAKLGGINQBLDMSIAEISPEEKERRKTMPLTMYGIDVTHPT-SYSGIDY-SI	732
Db	625	IVLKLMKMGINSRI-----VADKITNKYLVDPQTMVVGIDVTHPTQAEMRMNMPVS	677
Qy	733	RAVVASINPGGTIYRNMTVTOECPGERAVAHGERTDILEAKFVKLLREFAENNDNRA	792
Db	678	AAIVANVDLLPOSYGANVKVQKKCR---ESVYV---LLDAIRERIITFR-----HTKQK	726
Qy	793	PAHIVVYRDGVSDSMLRVSHDELRSLSKSEVKQFMSERDGEPEPKYTFIVIQKHNTRL	852
Db	727	PARIIIVYRDGVSEGFSEVLREEIQSIRTACLAI-----AEDFRPPIITVIVQGRHARI	781
Qy	853	LRMEKDKPVVKNKDLTPAETDVAVAAVKQWEEDMKESKETGIVNPSSGTTVDKLIIVSKYK	912
Db	782	CKYQNDMVGKAKNVPP-----	811
Qy	913	EDFELASHHGVLGTSRPGHYTYMYDDKGMSSODEVYKMTYGLAFLSARCKRPSLSPVPHY	972
Db	812	PDFYLCSHYGVQGSTSRPARYHVLDECKFTEADISQITTYGMCHTYGCRTRSISIPVYVY	871
Qy	973	AHLSCEKAK 981	
Db	872	ADLVATRAR 880	
RESULT 9			
D88568			
protein ZK757.3 [imported] - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004			
C:Accession: D88568			
R:Anonymous, The C. elegans Sequencing Consortium.			
Science 282, 2012-2018, 1998			
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo			
A:Reference number: A75000; MUID:99069613; PMID:9851916			
A>Note: see websites genome.wustl.edu/sec/C_elegans/ and www.sanger.ac.uk/projects/C.el			
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an			
A:Accession: D88568			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1040 <STO>			
A:Cross-references: UNIPROT:P34681; UNIPARC:UPI000013BC28; GB:chr_III; PIDN:CAA82941.1;			
C:Genetics:			
A:Gene: ZK757.3			
A:Map position: 3			
C:Superfamily: rabbit translation initiation factor eIF-2C			



Query Match	10.0%;	Score 537;	DB 2;	Length 1040;
Best Local Similarity	21.6%;	Pred. No. 2.9e-22;		
Matches 222;	Conservative 170;	Mismatches 375;	Indels 262;	Gaps 39;
Qy	45	NWKFSSKIYDREYVEYVMTKEVLRNPKGKPPKTEIPDRAKLFWQ-----HLR 98		
Db	104	NFFAMDLPKQKVVQYHVEI-----HHPGCKRLDKDEMRI-----IFWKA VSDHNPFI 152		
Qy	99	HEKQTDPILEDYVDEKDTVYSVCRLL---NTVTSKMLVSEKVKVQSEKKDEKLEKKI 155		
Db	153	HNK-----PAL---AYDGAHQLYTVARLEFPDQGSVRLDCASLPKONDRTRCAISIQN 205		
Qy	156	LYTMILTYRKXKHLNFSRENPEKDEEANRSYKFLKNVMTQKRVYAPFVNEBIKVQFAKNF 215		
Db	206	VGPVLEEMQR-----TRTN- NLDRELVTPQILDIIICRQSLT-CPLLKN-----SANP 251		
Qy	216	VYDNNLSLRVPESFHDNPRFQSLSEVAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFY 275		
Db	252	YTWKSSCYRIPTA-----AGQALDLBGKEMMTGFFSSAHIASNYRPLLNIDVAHTAFY 305		
Qy	276	NAPKMSLLDYLILLIVD-----PQ 293		
Db	306	KT-RITVLQFMCVLDVNETSKPENNPRPGPGPGGPGYGRGGRGGSGYGNFGNRP 364		
Qy	294	SCN-----DDVRKD-----LKTMLMAGK---MTIRQAAARPIRQLLENLKL 331		
Db	365	GANVRDDFGGNGLTFTMDTLSRDTQLSSFETRIFGDALRGKIRAAHRP----- 413		
Qy	332	KCAEVWNEMSRLTERHLFTLDLCEENS LVYKVTGKSDRGNAKKYDITTLFKIYEENKVF 391		
Db	414	NAIRVYKVNLSQLPADKLMFGQIDEEGRQV--VGSVAD-----YFSEKYGP 457		
Qy	392	IEPPLHLPLVKVSKAKAYAVPMHELVHEHKPQRYKNRIDLVMQDKFLKBATRKPHDYKEN 451		
Db	458	LKYEPQLCHLVGPPTNRIFLPMEHCLT--DSPQKYNKKMSKQTSAILIKAADVATQREDR 516		
Qy	452	TLKMLKBLDFSSEELNFERFGLCSQLMTECPKVLKEPMLVNSVNEQIKMTPVIRGFQ 511		
Db	517	IKQLAAQASFGTDP--FLKEFGVAVSQMIQTARVLIQPP-----PIMGGN 561		
Qy	512	EKQLVNVPEKELCAVP-----VNETAGNP--CLEENDV---KFYTELIGGCK-- 556		
Db	562	NRSVNPV-----VFPKDGSTMDNQTLMPATCRSYSMLALVDPDQTSLTQTPCQSL 613		
Qy	557	-FRGIRIGANENRCAQSIWYDATKNEYAFYKNCITLNTIG- RPEIATEAKNWERLPDX 614		
Db	614	TMKATAMGMPRPMPDLVKYGRSKBDV-----CTLFTFIADEVRVTMTVCD----- 659		
Qy	615	EQVLMFIITISKQLNAYGFVGHYCDHTIGVANHITSETVTYKALASLRHEKSKRIFYQ 674		
Db	660	-----CIIVVLQSKNSDIYMTVKEQSDIVHGIMSQCVLMKNVSRP-----TPATCAN 706		
Qy	675	IALKINAKLGGINQELDWSBEIAETISPEKERRKTMPLTYMVGIDVTHPT--SYSGIDY-SI 732		
Db	707	IVLKLANKMGGINSR-----VADKIITNKYLVQDPTWVTVGIDVTHPTQAEEMGNMPSV 759		
Qy	733	AAVVASINPGGTIYRNWIVTQEECPERAVAGHRETDILEAKFKVLKLEFPAENNDRA 792		
Db	760	AAIVAVDILLPQSYGANVKVQKKCR-----ESVYV---LLDAIRERIITFYR-----HTKQK 808		
Qy	793	PAHIVVYRDGVSDEMLRVSHDELSLKSEVKQPMSEDRGEDDEPKYTFETIVIOKRNTRL 852		
Db	809	PARIIIVRDGVSEGFSEVLRBEELQSI RTACLAI-----AEDFRPPTIYIVQKRHHARI 863		
Qy	853	LRRMEKD KPVVYNKDLTPAETDVA VAAVKQWEEDMKESKETGI VNPSSGTTTVDKLLIVSKYK 912		
Db	864	PKCYQNDMWGKAKNVP-----GTTVDGTGIVSPEG 893		
Qy	913	FDFPLASHGVLTGTSRGHTVTMYDDDKMSQDEVYKVTYGLAFSLASRCRKPISLPVPVHY 972		
Db	894	FDFVLCSHYGVQGTSRPARYHVLIDDECKFADEIQTISITYGCHTYGRTGSVSIPTPVY 953		

[illegible]



Db 609 ---FVDNLIDRCWRLGWMEA-----PIVYKTSRME-----TLSNGNAIEEL---- 647  
QY 604 AKMFERLPDKE---OKVLMFIIISKRQLNAGFVKHYCDHWIGVANQHITSETVTYKALA 660  
Db 648 LRSVIDEASRKHGGARPTIIVLCAMSRKD-DGYKTLKWIATKLGVLVTCQFLTPAT---- 702  
QY 661 SLRHEKSGRIFFQIATKINAKLGINQELDSEIAEISPEEKERRKTMPLTWYVGIDVT 720  
Db 703 -----KGGQYRANLALONAKVGSNVEL---MDTSPFFKKEDE-----VNFIGADVN 748  
QY 721 HPTSYSGIDYSIAAVASIN-PGGTIYRNMIVTQEECRPGERAVAHGRETDILEAKFYK 779  
Db 749 HPAARDKMSPSIVAVVGTLTWPEANRYAARVIAQ-----PHRKEEIQGFGDACLE 798  
QY 780 LLREFAENNDNAPAHIVYRQGVSDSEMLRVSHDELRLSKSEVKQFMSEKEDGEPKY 839  
Db 799 LVKAHVQATGQR-PNKIVIFRDGVSDAQDFMLNVLELDVK-----LTFEKNGYN--PKI 850  
QY 840 TFIIVQKRNTRLRLRMKDKPVNKKDLTPAETDVAVAAVKQWEDMKESKETGIWNPSS 899  
Db 851 TVIVAQRHQTRFF-----PATNND-----GSDKG--NVPS 879  
QY 900 GTTVDKLIYSKYKFPDFFLASHHGVLTSPGHYTYMYDDKMGSDQEVYKMTYGLAFLSAR 959  
Db 880 GTVWDTKVHPYEDFYLCSHHGGIGTSKPTHYTLMDLDELGFTSDQVQKLIFEMCFTR 939  
QY 960 CRKPTSLPVVHYAHLSCAKELRYTYKEHYIGDYAQR 999  
Db 940 CTKPVSLVPVYADVAFRGRMYHEASSRE--KNFKQPR 977

RESULT 11  
T52134  
Zwille protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004  
C:Accession: T52134  
R:Laux, T.  
submitted to the EMBL Data Library, January 1998  
A:Reference number: Z25975  
A:Accession: T52134  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-988 <LAU>  
A:Cross-references: UNIPROT:Q9XGW1; UNIPARC:UPI000016DA03; EMBL:AJ223508; PIDN:CAAL14239.  
C:Genetics:  
A:Gene: zwille  
A:Map position: 5  
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.1%; Score 488; DB 2; Length 988;  
Best Local Similarity 22.7%; Pred. No. 1.5e-19;  
Matches 221; Conservative 155; Mismatches 377; Indels 222; Gaps 38;  
QY 49 FSSKYDREYVEYEVQMTKEVLRNPKGPPKTEIPIDRAKLFQHLRHEKKOTDFTL 108  
Db 148 FIADPTKDLQNDVTITFEVSKSVNRAI-----IAELVRLY-----KESDLGR 192  
QY 109 EDYVFEKDTVYVSVCLNTVTSKMLVSEKVKKD-----SEKDEKDKLEKKILYTMILTVR 164  
Db 193 RLPAYDGRKSLVTAGELPPTWKEF--SVKIVDEDDGIINGPKRERSYKVAIKF---VARA 247  
QY 165 KKFHLN--PSRENPEKDEANRSYKFLKNVMTQKRYAP-----FVNEEIKVQFKNFYVD 218  
Db 248 NMHHLGEFLAGRADCPQBAVQILDIVLRELSVK-RFCVPGRSFFSPDIKT----- 297  
QY 219 NNSILRVPSFHDNRFNRFOSLEVAPRIEAWFGIYIGIKELFDGEPVLNPAIVDKLFYN-A 277  
Db 298 -----PQLSGEL-----ESWCGFYQSIRPTQGLS-LNIDMASAAFIPL 337  
QY 278 PKMSLLDYLL---LIVDPQSCNDVVRKDKLTKMLAKGKMTIROAARPRIRQLLENLKLKA 334  
Db 338 PVIEFVAQLGKDVLSKPLSDSRVK--IKKGLRGVKVEVTHRANVRRKYRVAGLTTQ-- 393

QY 335 EYWDNEMSLTBRHLTELDLCEENSLVYKVTGKSDRGRNAKKYDTTLFKIYBENKKF-IE 393  
Db 394 -----PTRELMP--PVDEN-----CTWKS-----VIEFYQEMYGFTIQ 424  
QY 394 FPHLPVLYKSGAKGYAVPMHELVHEKPKQRYKNRIDLYMODKFLKRAKTRKPHDYKENTL 453  
Db 425 HTHLPCLQVGNQKASYLPMACKIVE-QQRYTKRLNEKQITALLKVTQRAEGQRNDIL 483  
QY 454 KMLKELDFSSSELNFERPGLCSKLQMIKCPKVLKEPMLVNSVNEQIK-MTPVI---RG 509  
Db 484 RTVQHNVDQDP--YAKEFGMNISEKLASVEARILPAPWLKYHENGKEDCLPQVQGNM 541  
QY 510 FOEKOLNVVPEKELCCAVFVNNETAGNCPCLBENDVVKFYTELIGGCKFKGIRIGANENRG 569  
Db 542 MNKMGINGMTVSRWACVNF-----SRVQENVARGFQCNELQOMCEVSEMFNPFPVIP 594  
QY 570 AQSIIMYDATKNEYAFYKNCNLNTIGRFEIAATEAKNMFERLPDKQKQVLMFIIISKRQL 639  
Db 595 IYSARPDQVEKALKHVYHTSMNKTGK-----ELELLALLPDNNGSL----- 637  
QY 630 NAYGFVKHYCDHTIGVANQHITSETVTYKALASLRHEKSGKRIFYQIATKINAKLGINQE 689  
Db 638 --YGLDKRICETELGLISQCLTKHFVKI-----SKQYLADVSLKINVMQGRNTV 686  
QY 690 LDMSIAEISPEEKERRKTMPL-----TWYVGIDVTHTPTSYSGIDYSIAAVASIN-PGG 743  
Db 687 L-----VDALSCR-----IPLVSDIPTIIFGADVTHPENGRESSPSIAAVASQWPEV 735  
QY 744 TIYRNMIVTQEECRPGERAVAHGRE-----RTDILEAKFVKLLREFAEENNDN 790  
Db 736 TKYAGLVCAQ-----AHRQELIQDLYKTQWDPVRGTVSGGMIRDLLISFRKATQG 785  
QY 791 RAPAHIVYRQGVSDSEMLRVSHDELRLSKSEVKQFMSEKEDGEPKY-----TFIVIOK 846  
Db 786 K-PLRIIFYRDGVSGQFYVLLYELDAIRKACASL-----EPNYQPPVTFIVVQK 835  
QY 847 RHNTLLRRMEKDKPVNKKDLTPAETDVAVAAVKQWEDMKESKETGIWNPSSGTTVDKL 906  
Db 836 RHHTRLPANNHRDK-----NSTDRSGNLLP--GTVVDYK 867  
QY 907 IVSKYKPDFFLASHHGVLTSPGHYTYMYDDKMGSDQEVYKMTYGLAFLSARCKPISL 966  
Db 868 ICHPTEPDFYLCSHAGIQGTSRPAHYVLMWENNFTADGIQSLTNNLCYTYARCTRSVSI 927  
QY 967 PVPVHYAHLSCAK 981  
Db 928 VPPAYYAHLAARFR 942

RESULT 12  
T41568  
argonaute-like protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T41568  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z21991  
A:Accession: T41568  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-834 <WOO>  
A:Cross-references: UNIPROT:O74957; UNIPARC:UPI000012D2A7; EMBL:AL023705; PIDN:CAAL19275  
A:Experimental source: strain 972h-; cosmid c736  
C:Genetics:  
A:Gene: SPDB:SPCC736.11  
A:Map position: 3  
A:Introns: 43/3  
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 9.0%; Score 484; DB 2; Length 834;  
Best Local Similarity 21.8%; Pred. No. 2e-19;

Matches	212;	Conservative	165;	Mismatches	372;	Indels	222;	Gaps	39;
Qy	38	KKVLLVNNFKPSSK	IYDREYYEYEVKMTKEVLNR	-----RPGKPPFKTBIPIPDRAKL	92				
Db	20	KQITLKANFFQI	-----ISLPNETINQYHIVGDGSRVPRK	-----QSOL	59				
Qy	93	FWOHLRHEKKQTD	FILEDVYFDEKDTVYVSVCLNTVTSKMLVSEKVKVKDSEKKDEKOLE	152					
Db	60	IW-NSKEVKQVFGSSWMSNV	YDGRSMCMWSKGDADGTIKVNIAGESHPREIFSIQKS-S	117					
Qy	153	KKILYTMILATYRKPHLNF	SRENPEKDEANRSYKFLKNVNTQKRVYAPFVNEEBIKVOFA	212					
Db	118	KINLHTLSQFVNSKY	-----SSDPQVLSSIMFLDLLKKK	-----PSETLPGFM	161				
Qy	213	KNFYVDNNSILRVSPESHDP	PNRFOSLEVAPRIEAWFGIYIGIKELFDGPEVPLNPAIVDK	272					
Db	162	HSFPTGNGV	-----SLGGGEAWKGFYQSIIRP-NQGFNSVANDISS	203					
Qy	273	LFYNAPKMSLLDYLLLI	VDPOSCHNDVVRKDLTKLMAGKMTIRQAAPRIROLLENLKLK	332					
Db	204	AFWR--NDSLLOILMEYTD	CNSVRDLTRFDLK-----RLSKRFRFLKVT	245					
Qy	333	CAEYVDNEMSLTRHHTL	FFDLCEBNSIYVYKVTGSD-----RGRNAKKYDITTLFK	383					
Db	246	C-----QHRNNVGTDLANR	--VYSIEGFSSKSDSPFVRLNGEEQKISVAE	291					
Qy	384	IYEENKCP-IEFPHLPLV	KVKSAGKAYAVPMHEHLEVHEKPORYKNRDLVMQDFELKRAT	442					
Db	292	YFLENHNVRLOYPNLPC	ILVKNGA--MLPIEFCEV-VKGORIYAKLNSDQTANMIRPAV	347					
Qy	443	RKPHDYKENTLMLKEL	DFSSEELNFVERFGLCKLOMIECPGKVLKEPML--VNSVNE	499					
Db	348	QRPERVQOIDD	FVHQMDWDTP--YLTOYGMKIQKMLEVPARVLETPSPRISYGGDCIER	405					
Qy	500	QIKMTPIVRGQEKQLNV	PEKEL--CCAFVYVNTAGNCPLEBNDVVKFYTELIGCGCKFR	558					
Db	406	PVSGRWNLRG--	KRFLLDPPRAPIRSWAVMCFTSTRRLPMRGIEENFLQTVYQTL-----456						
Qy	559	GIRIGANENRGAQSI	MDATKNEYAFYKNCITLNTGIGRFEIAAT--EAKNMFERLPDKE	615					
Db	457	TSUGINFWMKPP	VIADIRGSVE-----ELCITLYKKAQOVGNAPD--499						
Qy	616	QKVLMPITIIISKRQL	NAYGFVKHYCDHTIGVANQHITSETVTKALASLRHEKSGKRIFQYI	675					
Db	500	--YLFILDKN	SPYGSIKRVCTNMLGVPSQCAISKHILQS-----KPQYCANL	547					
Qy	676	ALKINAKLGGINQEL	DSETABISPEEKERRKTWPL-----TMVYGVIDVTHP-TSYSIDY	730					
Db	548	GKMINVKGGIN	-----CSLIP-----KSNPLGNVPTLIILGGDVTHPGVGATGV--591						
Qy	731	SIAAVASINPGGTI	YRNMIVTQSECRPGERAVAHGRERTDILEAKFVCLLRFEAENNDN	790					
Db	592	SIASIVASVDLNG	KYTAVERSQ--PRHQEVIEGK--DIV-----VYLQGFRAWT-K	640					
Qy	791	RAPAHIVVYRDGVSD	SEMLRVSHDELRLSKSEVKQFMSERDGEDPBPKYTFIVIQKRHNT	850					
Db	641	QOPORIIVFRDGT	SEGQFSLVINDELSQLKEACHSLSPKYN-----PKILVCTTQKRHHA	695					
Qy	851	RLLRMEKDPVVKNDLT	PAETDVAAVAKQWEEDMKESKETGIWNSSGTTVDKLVSK	910					
Db	696	RFP-----IKNK	-----SDGDRNG--NFLPGTITIEKHVTHP	724					
Qy	911	YKDFDFLASHHGVL	GTSRPGHYTVMYDDKMGMSQDEVYKMTYGLAFLSARCKPISLPPVP	970					
Db	725	YOYDFYLLSHPS	LOGSVFVHYTVLHDEIQMPDPQFQTLCNLCYVVARATSVAISLVPVP	784					
Qy	971	HYAHLSEKAK	981						
Db	785	YTAHLVSNLAR	795						

RESULT 13  
T25164

hypothetical protein T23D8.7 - *Caenorhabditis elegans*  
C/Species: *Caenorhabditis elegans*  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T25164  
E:Wild, A.  
submitted to the EMBL Data Library, October 1996  
A/Reference number: Z19989  
A/Accession: T25164  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-898 <WIL>  
A/Cross-references: UNIPROT:Q9XVI3; UNIPARC:UPI000007743E; EMBL:Z81128; PIDN:CA  
A/Experimental source: clone T23D8  
C/Genetics:  
A/Gene: CESP:T23D8.7  
A/Map position: 1  
A/Introns: 40/3; 63/2; 188/2; 508/3; 552/2; 785/3; 849/3  
C/Superfamily: rabbit translation initiation factor eIF-2C

Query Match	8.8%	Score 473;	DB 2;	Length 898;
Best Local Similarity	22.7%	Pred. No. 9.3e-19;		
Matches 242;	Conservative 160;	Mismatches 366;	Indels 296;	Gaps 46

Qy	2	SSNPPELEKGF-----YRSLDPE---MKWLARPTCKDCKGFVEKKVLLLVN	45
Db	31	SINVPSLENBFLSSSSGSRVSDLLYLHPIBEENRPFPLIGKLPLSTTGRF-----LSLIAN	86
Qy	46	WFKF--SSKLVDRYYEYEVKMTKEVLRNPKGPPKKTEIPIPDRAKLFQHLRHEKKQ	103
Db	87	HQIQTNGSIHQYIRDFDPIPSKUNR-----TIURTQEQ	124
Qy	104	TDFILE-DYVFDEKOTVYSVCLNTV---TSKMLVSEKVKVKDSEKDEKOLEKKILYTM	159
Db	125	NPGLIECLPFDGHTVYSTELINVKVNSVINAGVWNTKESPN-----LPKL	174
Qy	160	ILTYRKKFHLNFSRENPEKDEANRSYKPLKNVMTQKVRYPVNEBIKVQFAGNFVYDN	219
Db	175	YLTHVDSFLDTKIITGNQDQ-----NOKLRMHAIDTVFROTSTGNF-----	217
Qy	220	NSILRVSEPHDPNRFQS-----LEVAPRIEAWFGIYIGIKELFDGEPVLFAIVDK	272
Db	218	HAVLOSFFSAQNSAIEPSHGLGWGTYNLGVREVCYGFQNVVETFD-TLTMNLDVATT	276
Qy	273	LFNAPKMSLLDY-----LLIIVDQSGNDDVRKDLTKYLMAGKWTITQOARPIRQ	324
Db	277	TFYR--PVALVEPLAEILLEVPLATVTDGRSLSDVQKKKFNREVAGLVKVTREHCSPPRRF	334
Qy	325	L-----LENLKLKCAEVDNEMSRLTERHFLFDLCEBNSLYVKVTKGSDRGRNAKCY	377
Db	335	VARCTWKPTENISFHLSSETAGNQSCKP-----LSLVE---YK-----RRY	372
Qy	378	DTTLFKIYBENKGFIEPPHPLPVKVKSGAKEYAVPMMEHLEVHEHKPQRYQNRIDLVMDKGF	437
Db	373	NIDL-----TVKHLPCIEV-GRTRCITLPLELCYV-VSGQCIKKLN-EQOLIAN	418
Qy	438	LKBAIRKPHDYKENTLKMELDSESEELNFWRRFGLCSKLOMIECKGVLEKPMLV---	494
Db	419	LIRATSNATERQNAVMSLQNLKMDNDVNAV-KFGLKVEAQLKIEGRVLPVPRLLYRS	477
Qy	495	-NSVNEQIKWTP-----VIRGFQSKQLNVV---PE---KELCCAVFVN-----	531
Db	478	PNLKRQECVTPNNGTWMRGKNFYSGIQIREWAVCFASPEIIGEASMSFVZNLVNA	537
Qy	532	ETGAGNPLENDVVVFTELIIGGCKPFGIRIGANENRGAOSIMYDATKYNEVAFYNGCTLN	591
Db	538	SBIGMPFLEH---RF-----CRY-----	553
Qy	592	TGIGRPEIAATEAKNPFERLPDKEQKVLMEFIISKROLNAVGFVKHYCDHTIGVANOHIT	651
Db	554	-----AEPDQTVKLEHLENEQNLQVLICI VPGKSV-VIGELKVK-GBLLGLTTCQVR	604
Qy	652	SETVTKALASLRHEKSGSKRIFYOIALKINAKLGGINOELDWSEIABTSPEKEKRRKTMPL	711

```
Db 605 SQNVSKA-----SPHTLSNLCKMINKSLGGINVL-----SPQSLNSEPV-- 646
Qy 712 TMYGIDVTH-----PTSYSGIDYSIAAVASINPGGTYIRNMIVTQECRPGERA 762
Db 647 -LFIGCHLTRSLASSDSTSIAHCDSSIACLVSGMDGHPQFQSPIFRTQPR----- 698
Qy 763 VAHGRERTDILAKFVKLLRPAEN---NDNRPAHAHVYVRDGVSDSEMLRVSHDELRS 819
Db 699 --HQTIVDMCE-----MTREALINFRKSTGPKPKHIIYRAGIADVTVDEIMQTELRAV 751
Qy 820 KSEVQFMSERDGEDPEPKYPTIVIOKRNTRLRLRMKDKPVPVNDLTPASTDVAAVAV 879
Db 752 RDACA--MIEYGF--QPGITFGLDVTHTLFLFAANEKDR--VGNSSQNPAGTLV----- 800
Qy 880 KQWEDMKESKETGIVNPSGGTVDKLVSKYKFPDFFLASHHGVLTGSRPGHYTMYDDK 939
Db 801 -----BTGI-----TVNVL-----FEFYLVSAGIOGTSRPTKYVVMWDDN 836
Qy 940 GMSODEVYKMTYGLAFLSARCKRPISLPVPVHYAHLSCAKEL 983
Db 837 SIPSADIHEMTYQLCHTQSRCTRSVSPSPVYAKLVQAQAKIL 880

RESULT 14
A:Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: AB4668
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Status: preliminary
A:Accession: AB4668
A:Molecule type: DNA
A:Residues: 1-930 <STO>
A:Cross-references: UNIPROT:Q9ZVD5; UNIPARC:UPI0000179101; GB:AE002093; NID:g3885334; PI
C:Genetics:
A:Gene: At2g27040
A:Map position: 2
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 8.0%; Score 428.5; DB 2; Length 930;
Best Local Similarity 22.0%; Pred. No. 3.le-16;
Matches 218; Conservative 167; Mismatches 381; Indels 223; Gaps 40;

Qy 38 KKVLLVNMFKSSKIYDREYVEYVMTKEVLNRKPKGPKPKK-TEIPIDRAKLFWQH 96
Db 67 QKIPLLTWHFKVDVANLQGHFFHYVALFYD-----DGRPVEQKGVRKILDKV----- 115
Qy 97 LRHEKKQTDFILEDYVDEKOTVYS-----VCLRLNTVTSKMLVSE----- 136
Db 116 --HQTVSHSLDCKEFAYDGEKLTFTYGALPSNKMDPSVLEBVSATSKDFVSRRANGNSP 173
Qy 137 --KVYKSEKKDQKLEKILYTMILTYRKFFHLNFSRENPEKDEEARSFKPLKNVMT 194
Db 174 NGNESPSPGDRKRLRPNRSKNFRVSIYAAKIPLQ-ALANAMRGQESNSQEAIR----- 228
Qy 195 OKRVYAPFVNEIKVQAKNFYVDNNSILRVPESE--HDPNFEQSLVAPPIEAWFGLY 252
Db 229 -----VLDIILRQHA-----RQGLLVRSQSFHNDPTNCE---PVGGNILGCRGFH 272
Qy 253 IGKELFGPEPVNFAIDKLFYNAPKMSLLDYLIVDPOSCNDVDVRKDLTKLMAGKM 312
Db 273 SSFRIT-QQGMSLANDVTTMIKFG--PVVDFLI-----ANQNARDPYSIDWSKAKR 322
Qy 313 TIROAARPIROLLNLKLCAEVDN-EMSLRTEHLTFLDLCENSLVYKVKTSGRG 371
Db 323 T-----LKNLRVKVSPSQGEFKITGLSDPK-----CREQTFELKKNPNENG 364
```

```
Qy 372 RNAKYDTTLFKI--YEENKKFIEPPH---LPLVKVSGAKAYAVPMHELVHEKQRYK 426
Db 365 ----EFETTEVTADYFRDTHIDLOYADLPFCINVGKPKRTYIPLE-LCALVPLQRYT 419
Qy 427 NRIDLVMQDKFLKRAKTRKPHDYKENTLQKLEDFSSBELNVPFRGLCSKLIQMIKCPK 486
Db 420 KALTTFORSALVERSKRQPKQBRMTVLSKALKVSNYDABPL--LRSCGISISSNFTQVEGR 477
Qy 487 VLKEPMLVSNVEQIKMTFVIRGFOEKQLNVVPEKELCCAVFVNNETAG-NFCLSENDVV 545
Db 478 VLPAPKLKMGCGSETFPRNGRWNFNKKEF-VEPTK--TORVVVVPFSAKCNVRQVVDLLI 534
Qy 546 KFYTELIGCGKFRGRIGA-----NENRGAQSIMYDATKNEYAFYKKNCTLNTGIGRF 597
Db 535 K-----IGSK--GLEIASPPQVPEEGQFRAPPWI-----564
Qy 598 EIAATEAKMFERLPDKQKVLMTFII---SKQLNAYGFVKHYCDHTGTIGVANQHITSET 654
Db 565 ----RVENMPKDIQSKLPGVPOFILCVLPDKKNSDLYGPMWKKQLNTEFGI----- 610
Qy 655 VTKALASLRHEKSKRIFYQIALKINAKLGGINQELDWSEIAEISPEEKERRKTMPLTWY 714
Db 611 VTQCHAPTR--QPNDQYLTNLLKINAKLGGINSLSVERTPAFTVISK-----VP-TII 662
Qy 715 VGIDVTPTSYSGIDYSIAAVASIN--PGGTYIRNMIVTQECRCPGERAV-----AHGR 767
Db 663 LGMDVSHGSPQSDVPSIAAVSSREWPLISKYRASVTRQ-----PSKAEMIESLVKNGT 718
Qy 768 ERTDILEAKFVKLLRFEANNDNRPAHIVVYRQGVSDSEMLRVSHDELRSKSEVKQPM 827
Db 719 EDDGIK----ELLAVDFYTSNNKRKPEHIIIPRDGVSESQFNQVLNIELDQIIEACKLLD 774
Qy 828 SERGEDPEPKYTFYVIOKRNTRLRLRMKDKPVPVNDLTPAETDVAVAVKQWEEDMK 887
Db 775 ANWN-----PKPFLLVAKQNHHTKFP-----QPTSPENVP-----805
Qy 888 ESKETGIVNPSGGTVDKLVSKYKFPDFFLASHHGVLTGSRPGHYTMYDDKGMQDEYV 947
Db 806 -----GTIIDNKICHPKNDYFLCAHAGMIGTTRPTHYHVLVDEIGFSADEIQ 853
Qy 948 KMTYGLAFLSARCKRPISLPVPVHYAHL 976
Db 854 ELVHSLSYVYQRTSASISVAPICVAHLA 882

RESULT 15
T01113
Translation initiation factor eIF-2C homolog T21L14.12 - Arabidopsis thaliana
N:Alternate names: Argonaute (AGO1)-like protein
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: T01113; D84739
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau-
submitted to the EMBL Data Library, December 1997
A:Description: Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence.
A:Reference number: Z14209
A:Accession: T01113
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-887 <ROU>
A:Cross-references: UNIPROT:Q48771; UNIPARC:UPI00000AB808; EMBL:AC003033; NID:g2702261;
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Status: preliminary
A:Accession: D84739
A:Molecule type: DNA
A:Residues: 1-887 <STO>
A:Cross-references: UNIPARC:UPI00000AB808; GB:AE002093; NID:g2702284; PIDN:AA91987.1; (
```

## C:\Genetics:

A:Gene: AC2932940; T21L14.12

A:Map position: 2

A:Introns: 61/3; 130/2; 203/2; 248/1; 281/3; 314/2; 367/3; 406/3; 442/3; 463/3; 501/3; 5

C:Superfamily: rabbit translation initiation factor eIF-2C

## Query Match

Best Local Similarity 22.4%; Pred. No. 8.8e-15;

Matches 227; Conservative 160; Mismatches 358; Indels 268; Gaps 45;

Query Match 7.58; Score 402; DB 2; Length 887;

Qy 939 KGMQSDDEVYKMTYGLAF-----LSARCRKPISLPVPHYAHLSCSEKAKE 982  
| | | | | : : : : : | | | | | : : : : :  
Db 806 IGSPDDQLNLHLSYKLNLSIFNVSSLLCVFLSV-APVRYAHLAAQVAQ 857

Search completed: July 5, 2006, 13:34:44  
Job time : 56 secs

Qy 42 LLVNWKFSSKIYDREYIEYEVKRLKPKPPKTEIPDRAKLFWQHLRHEK 101  
| | | | | : : : : : | | | | | : : : : :  
Db 41 LCTNHFNVSVRQPDVVVFYQYTVSITTEGDAVDGTGSRKL-----MDQLP-----KT 88

Qy 102 KQTDFILEDVDEKOTVYVCRK--NTVTSKMLVSEKVVKDKSEKDE-----KOLE 152  
| | | | | : : : : : | | | | | : : : : :  
Db 89 YSDDLGRKLAYDGEKTLTYVGPLQNEFDLVIVGFSFGSKDGVSDGSSSGTCKRSK 148

Qy 153 KKILYTMILTYRKKFHLNFSRENPEKDEEANSYKFLKNYM-TQKVRYAPFVNEEIKVQF 211  
| | | | | : : : : : | | | | | : : : : :  
Db 149 RSLPR---SYKQIIRH---YAAEIP-----LKTVLGTORGAYTPDKSAQ----- 186

Qy 212 AKNFVYDNNLSILRVPSFHDPNRFEQSLVAPRIEAMFGIYIGIKELF-----DGE 263  
| | | | | : : : : : | | | | | : : : : :  
Db 187 -----DALRV-----LDIVLQQAARCGCLLVQAFFHSDGHPMKVGG 224

Qy 264 -VNFALVDKLP---YNAPKMSLLDYLLIVDPQSCNDVDRKDLTKLMAGKMTIRQAAR 319  
| | | | | : : : : : | | | | | : : : : :  
Db 225 GVIGIRGLHSSFRPTGGSLNLDVSTTMILEPGVPVIEFLKANQSVETPRQIDWIKVAAK 284

Qy 320 PRIROLLENLKLKCAEWDNEMSLRTERHTFLDL-----CHENSILVYKVKTSRGRN 373  
| | | | | : : : : : | | | | | : : : : :  
Db 285 -----MLKMRVKA-----THRNMEFKIIGLSSKPCNQQLFSMKI-----KDG 325

Qy 374 AKKYDTTLTKIYEEN-KKFTFPHPLVVKYGAKEVAPME-----HLEVHEKPKQYKN 427  
| | | | | : : : : : | | | | | : : : : :  
Db 326 VPIREITVDYDFKQYTEPISSAYFPCLDVKGKDRPNYLPLEFCNLVSLQRYTKPLSGRQ 385

Qy 428 RIDLVNQDKPLKRAKTRPHDYKENTLKMKLKD-----FSSEELNFVERFGLCSKLQMI 482  
| | | | | : : : : : | | | | | : : : : :  
Db 386 RVLLV-----ESSRQKP-----LERIKTLANDAMHTYCYDKDPFLAGGGSIEKEMTQ 432

Qy 483 CPCKVLKEPMLVNSVNEQIKMTPIRGFQEKQLNVVPEKELCCAVFVNETAGNPLEEN 542  
| | | | | : : : : : | | | | | : : : : :  
Db 433 VEGRVLPKPMLEKFGKNEFQPCNGRWNFNKML---LEPRAIKSWAIVNFSF--PC--- 483

Qy 543 DVVKFYTELIGGCKFGIRIGANENRGAOSIMYDATKNEYAFYKNCNTLNTGIGRPEIAAT 602  
| | | | | : : : : : | | | | | : : : : :  
Db 484 DSHISRELI-SC---GMRKGIEDRPALVEEDPQ-----YKKA-----GPVERVEK 527

Qy 603 EAKNMPERLPDKEQKVLPMFIILSKROLNAYGFVKHYCDHTIGVANQHIITSETV-TKALAS 661  
| | | | | : : : : : | | | | | : : : : :  
Db 528 MIATMKLKFPPDPHFILC-ILPERKTSDIYGPWKIC-----LTEEIGHTQCICP 576

Qy 662 LRHEKSKRIFYQIALKINAKLGINQELDWSIAISPEEKERRKTMPL-----TWYVG 716  
| | | | | : : : : : | | | | | : : : : :  
Db 577 I---KISDQVLTNVLKINSKLGINSLLG-----IEYSNIPLINKIPTLIIG 622

Qy 717 IDVTH-PTSYSGIDYSIAAVAS-INPGGTIYRNMIVTQ-----EECRPGE 760  
| | | | | : : : : : | | | | | : : : : :  
Db 623 MDVSHGPPGADVP-SVAAVGSKCWPLISRYRAAVRTQSPLEIMDSLPQPIENTEKGD 681

Qy 761 RAVAHGRERTDILEAKFVKLLREFANNDNRAPAHIVVYRDGVSDSEMLRVSHDELRSK 820  
| | | | | : : : : : | | | | | : : : : :  
Db 682 NGIMN-----ELFVEFYRTSRARKPKQIIIFRDGVSESQFEQV-----LK 721

Qy 821 SEVKQFMS---ERDGEDPEPKYTFIVIOKRNTLLRRMEKDKFPVVKDLTPAETDVAVAA 878  
| | | | | : : : : : | | | | | : : : : :  
Db 722 IEVDQIIKAYQRLGESDVPKFTVIVAQKNHTKLF----- 756

Qy 879 VKQWBEEDMKESKETGVNPSSTGTTVDKLIIVSKYKFDFFLASHHGVLTGSRPGHYTVMYDD 938  
| | | | | : : : : : | | | | | : : : : :  
Db 757 -----QAKGPENVPAGTVVDTKIVHPNTYDFYMCNHAGKIGTSRPAHYHVLDE 805

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM protein - protein search, using sw model

Run on: July 5, 2006, 13:25:16 ; Search time 315 Seconds  
(without alignments)  
2995.289 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 5349  
Sequence: 1 MSSNPFLEKGFYRHSIDPE.....RHEMFLOQTNVYKPGMSFA 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Uniprot 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5349	100.0	1020	Q9XU82 CAEEL	Q9x82 caenorhabdi
2	3241	60.6	976	Q60V50 CAEER	Q60v50 caenorhabdi
3	604.5	11.3	997	AGOL ARATH	Q9sjk3 arabidopsis
4	577	10.8	857	I2C1 HUMAN	Q9ul18 homo sapien
5	573	10.7	863	Q4RKH3 TETNG	Q4rk3 tetraodon n
6	571.5	10.7	848	Q8CGU0 MOUSE	Q8cg0 mus musculu
7	571	10.7	857	I2C1 MOUSE	Q8c9i mus musculu
8	568.5	10.6	1055	Q7Y001 ORYSA	Q7y001 oryza sativ
9	567.5	10.6	891	Q7P241 ANOGA	Q7p241 anopheles g
10	565.5	10.6	862	Q6CKX2 XENLA	Q6cx2 xenopus lae
11	565	10.6	840	I2C2 RABIT	Q77503 oryctolagus
12	565	10.6	860	I2C2 MOUSE	Q8c90 mus musculu
13	565	10.6	871	Q6DUB9 XENTR	Q6dj59 xenopus tro
14	564	10.5	860	I2C2 RAT	Q9qz81 rattus norv
15	564	10.5	883	Q4VAB3 MOUSE	Q4vab3 mus musculu
16	563	10.5	860	Q6QME8 BOVIN	Q6qme8 bos taurus
17	563	10.5	1024	Q61PV1 CAEER	Q61pv1 caenorhabdi
18	560.5	10.5	950	Q9V312 DROME	Q9v312 drosophila
19	560.5	10.5	984	Q9V6V6 DROSOP	Q9v6v6 drosophila
20	558	10.4	859	Q4S7L5 TETNG	Q4s7l5 tetraodon n
21	557	10.4	1002	Q20578 CAEEL	Q20578 caenorhabdi
22	557	10.4	1010	Q3LFR7 CAEEL	Q3lfr7 caenorhabdi
23	556.5	10.4	851	I2C2 HUMAN	Q9ukv8 homo sapien
24	555.5	10.4	703	Q571J6 MOUSE	Q571j6 mus musculu
25	555.5	10.4	860	I2C3 HUMAN	Q9h9g7 homo sapien
26	555.5	10.4	860	I2C3 MOUSE	Q8cjf9 mus musculu
27	555	10.4	861	I2C4 HUMAN	Q9hck5 homo sapien
28	555	10.4	861	Q5VXF0 HUMAN	Q5vxf0 homo sapien
29	554.5	10.4	782	Q5TA58 HUMAN	Q5ta58 homo sapien
30	553	10.3	861	Q4VBD7 MOUSE	Q4vbd7 mus musculu
31	553	10.3	951	Q8BTF4 MOUSE	Q8btf4 mus musculu

#### RESULT 1

Q9XU82 CAEEL  
ID Q9XU82 CAEEL PRELIMINARY; PRT; 1020 AA.  
AC Q9XU82; Q9U6Q1;  
DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2001, sequence version 2.  
DT 07-FEB-2006, entry version 26.  
DE Hypothetical protein rde-1 (RNA interference promoting factor RDE-1).  
GN Name=rde-1; ORFNames=K08H10.7;  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;  
RG The C. elegans sequencing consortium;  
RT "Genome sequence of the nematode C. elegans: a platform for investigating biology."  
RL Science 282:2012-2018(1998).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20004399; PubMed=10535731; DOI=10.1016/S0092-8674(00)81644-X;  
RA Tabara H., Sarkissian M., Kelly W.G., Flenor J., Grishok A.,  
RA Timmons L., Fire A., Mello C.C.;  
RT "The rde-1 gene, RNA interference, and transposon silencing in C. elegans."  
RL Cell 99:123-132(1999).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
EMBL; Z83113; CAB05546.2; -; Genomic\_DNA.  
EMBL; AF180730; AAF06159.1; -; mRNA.  
PIR; T23510; T23510.  
DR Ensembl; K08H10.7; Caenorhabditis elegans.  
DR WormBase; WBGene0004323; rde-1.  
DR WormPep; K08H10.7; CE28243.  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003185; Pwi.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Pwi; 1.  
DR PROSITE; P85082; Pwi; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 1020 AA, 118804 MW, 8E2F1A2EFC43A670 CRC64;

Query Match 100.0%; Score 5349; DB 2; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 1.2e-292;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNPFLEKGFYRHSIDPEMKWLARPTGKCDGKFKYKKVLLVNNFKSSKIYDREYYE 60  
DB 1 MSSNPFLEKGFYRHSIDPEMKWLARPTGKCDGKFKYKKVLLVNNFKSSKIYDREYYE 60

```
QY 61 YEVKMTKEVLNRPKPKTEIPIPDRAKLPWHLRHEKKQTDPILEDYVVDSEKQTVY 120
Db 61 YEVKMTKEVLNRPKPKTEIPIPDRAKLPWHLRHEKKQTDPILEDYVVDSEKQTVY 120
QY 121 SVCLNTVTSKMLVSEKVKVKKDSEKKDEKLEKKILYTMILTYRKKFHLNFSRENPEKDE 180
Db 121 SVCLNTVTSKMLVSEKVKVKKDSEKKDEKLEKKILYTMILTYRKKFHLNFSRENPEKDE 180
QY 181 EANSYKFLKNVMTQKRVAPFVNBEEKVQPAKNFYVDNNSILRVPESHDPNRFQESLE 240
Db 181 EANSYKFLKNVMTQKRVAPFVNBEEKVQPAKNFYVDNNSILRVPESHDPNRFQESLE 240
QY 241 VAPRIEAMFGIYIGIKELPDGEPVLNFAVDKLFYNAPKMSLLDYLILLVDPQSCNDDVR 300
Db 241 VAPRIEAMFGIYIGIKELPDGEPVLNFAVDKLFYNAPKMSLLDYLILLVDPQSCNDDVR 300
QY 301 KDLTKLMAGKMTIRQAARPRIRQLLENLKLCAEVDNEMSRLLTERHLTFDLDCEENSL 360
Db 301 KDLTKLMAGKMTIRQAARPRIRQLLENLKLCAEVDNEMSRLLTERHLTFDLDCEENSL 360
QY 361 VYKVGKSDRGNAKKYDTTLTKIYEENKFTIEFPHLPLVKYKGAKEYAVPMHEHLEVE 420
Db 361 VYKVGKSDRGNAKKYDTTLTKIYEENKFTIEFPHLPLVKYKGAKEYAVPMHEHLEVE 420
QY 421 KPQRYKNRIDLVMQDKFLKGRATKPDHYKENTLKMKELDPSSEELNFSVERFGLCSKQM 480
Db 421 KPQRYKNRIDLVMQDKFLKGRATKPDHYKENTLKMKELDPSSEELNFSVERFGLCSKQM 480
QY 481 IECPGKVLKEPMLVNSVNEQIKMTPIVIRGFBKQLNVNPEKELCCAVFVNNTAGNPCLE 540
Db 481 IECPGKVLKEPMLVNSVNEQIKMTPIVIRGFBKQLNVNPEKELCCAVFVNNTAGNPCLE 540
QY 541 ENDVVKFYTELIGCGKFRGIRGANENRGAQSIMYDATKNEYAFYKNCYTLNTGIRFEITA 600
Db 541 ENDVVKFYTELIGCGKFRGIRGANENRGAQSIMYDATKNEYAFYKNCYTLNTGIRFEITA 600
QY 601 ATEAKNMFERLPDKQKQVLMFIIISKRLNAYGFVKHYCDHTIGVANQHITSETVTYKALA 660
Db 601 ATEAKNMFERLPDKQKQVLMFIIISKRLNAYGFVKHYCDHTIGVANQHITSETVTYKALA 660
QY 661 SLRHEKSGRIYQIALKINAKLGGINQELDWSAEISPEEKERRKTMPLTMVYGVDT 720
Db 661 SLRHEKSGRIYQIALKINAKLGGINQELDWSAEISPEEKERRKTMPLTMVYGVDT 720
QY 721 HPTSYSGIDYSAAVVASINPGGTYIRNMIVTQECRPGERAVAGHRETDILEAKFVKL 780
Db 721 HPTSYSGIDYSAAVVASINPGGTYIRNMIVTQECRPGERAVAGHRETDILEAKFVKL 780
QY 781 LREPAENNDNRAPIHVVYRDGVSDSEMLRVSHDELRSLSKSVKQFMSRDEGDEPEPKYT 840
Db 781 LREPAENNDNRAPIHVVYRDGVSDSEMLRVSHDELRSLSKSVKQFMSRDEGDEPEPKYT 840
QY 841 FVIOQRHNTRLLRMEKDKPVVNKDLTPAETDVAVAVKQNEEDMKESKETGIYNPSSG 900
Db 841 FVIOQRHNTRLLRMEKDKPVVNKDLTPAETDVAVAVKQNEEDMKESKETGIYNPSSG 900
QY 901 TTVDKLVISKYKDFPLASHHGVLGTSRPGCHTYTMVDDKGMQDEYVYKTYGLAFLSARC 960
Db 901 TTVDKLVISKYKDFPLASHHGVLGTSRPGCHTYTMVDDKGMQDEYVYKTYGLAFLSARC 960
QY 961 RKPISIPVPHVYAHLSCEKAKELRYTYKEHYIGDVAQPRTRHEHFLQTNVYKPMQSPA 1020
Db 961 RKPISIPVPHVYAHLSCEKAKELRYTYKEHYIGDVAQPRTRHEHFLQTNVYKPMQSPA 1020
RESULT 2
Q60VS0_CAEBR PRELIMINARY; PRT; 976 AA.
AC Q60VS0;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 7.
```

```
DE Hypothetical protein CBG19426.
GN Name=CBG19426;
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI TaxID=6238;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=AF16;
RX PubMed=14624247; DOI=10.1371/journal.pbio.0000045;
RA Stein L.D., Bao Z., Blasiar D., Blumenthal T., Brent M.R., Chen N.,
RA Chinwalla A., Clarke L., Clee C., Coghlan A., Coulson A.,
RA D'Eustachio P., Fitch D.H.A., Fulton L.A., Fulton R.E.,
RA Griffiths-Jones S., Harris T.W., Hillier L.W., Kamath R.,
RA Kuwabara P.E., Mardis E.R., Marra M.A., Miner T.L., Minx P.,
RA Mullikin J.C., Plumb R.W., Rogers J., Schein J.E., Sohmann M.,
RA Spieth J., Stajich J.E., Wei C., Willey D., Wilson R.K., Durbin R.,
RA Waterston R.H.;
RT "The genome sequence of Caenorhabditis briggsae: a platform for
RT comparative genomics."
RL PLoS Biol. 1:166-192(2003).
CC -!- CAUTION: the sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAC01000098; CAB72296.1; -; Genomic_DNA.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; P1wi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; P1wi; 1.
DR PROSITE; PS50822; P1wi; 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 976 AA; 114611 MW; 8B5427451C9ADCAA CRC64;
SQ
Query Match 60.6%; Score 3241; DB 2; Length 976;
Best Local Similarity 62.7%; Pred. No. 7, 4e-174;
Matches 630; Conservative 136; Mismatches 206; Indels 32; Gaps 8;
QY 21 MKWLARPTGKCDGKPYEKKVLLVNMVFPSSKIYDREYVEYEVMTKE--VLNRKPKGKPF 78
Db 1 MKWLSRPEKCEGKYVEKVKLLVNMVFNNTIYDREYV-YFVDMYKCKYMDKRTKSPK 59
QY 79 PKKTEIPDRAKLPWHLRHEKKQTDPILEDYVVDSEKQTVYVCL-NTVTSKMLVSEK 137
Db 60 TKQMPYIPERSEAFWKHLRFEQKNPFQIEDYVDFDEKDTVYSIHKTPTNGMSSEI- 114
QY 138 VVKQDEKKDEKLEKKILYTMILTYRKKFHLNFSRENPEKDEEANSYKFLKNVMTQKV 197
Db 115 ---PDQNRDKLVLK-----MHYCKENLYFSREDPTRDEANRSCKFLKNVMTQKV 164
QY 198 RYAPFVNBEEKVQPAKNFYVDNNSILRVPESHDPNRFQESLEVPRIEAMFGIYIGKE 257
Db 165 RCAPYVANEIGQLAKNFYDGNLSILRVPESHDPDRFHSLSLEIAPRIESWFGIYIAVKE 224
QY 258 LFDGEPVLNFAVDKLFYNAPKMSLLDYLILLVDPQSCNDDVRKDLTKLMAGKMTIROA 317
Db 225 LFDGEPVLNFAIDKLFYNAPQMSLLDYLILLVDPQSHMDVRRNHKEQLNSGLNLTIKPA 284
QY 318 ARPRIRQLLENLKLCAEVDNEMSRLLTERHLTFDLDCEENSLVYKVTGKSDRGNAKKY 377
Db 285 MFQSIDQLLANLKLKAAASVWNPKRSEFAERHLTFIRLSKYNSSHEEMIPVPRDRDRNAPID 344
QY 378 DTTLFKIYEENKFTIEFPHLPLVKYKGAKEYAVPMHEHLEVEHVEKQRYKNRIDLVMQDKF 437
Db 345 RVPLFQIYEKRNKYIEFPRLPLVKCKSGNNEYSVPMFLEVYVEKQRYKNRIDFAMQDKF 404
QY 438 LKRATRKPDPHYKENTLKMKELDPSSEELNFSVERFGLCSKQWIECPGKVLKEPMLVNSV 497
Db 405 VNAATRDPMYKKEALEMKDLDFSSGSLNFSVERFGFTDLKMIKCVGKVLKEPMLVNSV 464
```



QY 498 NEOIKMTPTVIRGFOEKQALNVVPEKELCCAVVYNE-TAGNPCLEENDVVKFYTELIGGCK 556  
Db 465 NOKISMTFVVRGFOEKALNVVPEKELCCAVFLREDKGTPEPCLEEDVVRFYKTLIDGCE 524  
QY 557 FRGIRIGANENRGAQSIIMYDATKNEYAFYKNCITLNTIGIRFEIATAAETAKMFERLPDKEQ 616  
Db 525 FRSIRIGKHONSQVRSLLQDPESGKYGFYVNVQLTAGVRNFRACADAKAMFERLADKNN 584  
QY 617 KVLMPILISKRQLNAVGVFKHYCDHTIGVANQHITSETVTYKALASLRHEKSGKRIFYQIA 676  
Db 585 KILLFIVFSKRRWNFYGVKQFCDVELGVASQHVTAHVAKRALQDMSSHKPKSKRIFYQIA 644  
QY 677 LKINAKLGGINQELDSWETAIEISPEEKERKMTPLMTYTGIDVTHPTSYSGIDYSTAAVV 736  
Db 645 LKINGKLGGVQNELDSWENAEIMTEBEKKRKNPLRMYGVGIDVTHPTNGSGIDYSTAAIV 704  
QY 737 ASINPGGTYRNMIVTQEECRPERAVAHGRETDILEAKFVKLLREFAEENNDRAPAH 796  
Db 705 ASNPGGTYRNMIVTQEEENRPERPVANGRETDILEGKFKVLLRIFAENNENRIPSHI 764  
QY 797 VYTRGVSDSEMLRVSHDELRSLSKSEVQPMSEDEGPEPKYTFYVIOQRNTRLRLRM 856  
Db 765 VYTRGVSDSEMLRVSHDELRSLSKAEVERFLKERGNEQIPYTFIVLQKRHKTRLFRI 824  
QY 857 EKDKPVVKNKLTFAETDVAVAVKQHEEDKESKETGIVNPPSGTIVDKLIVSKYKEDPF 916  
Db 825 ENERP---KD-----EAAKRWDEELKASQNTGFVNPTSGTIVDRITVSKYKEDPF 872  
QY 917 LASHHGVLGTSRPGHYTVMYDDKMSODEVYKMTYGLAPLSARCRPISLPVPHVYAHLS 976  
Db 873 LASHHGALGTSRPGYTVVHDDKMTKDELYKMTYELSFLSARCRPISLPVPHVYAHLS 932  
QY 977 CEKAKELYRXYKEHYIGDYAQPRTHEMEHFLQTNVVKYPMSPA 1020  
Db 933 CEKAKELYRCKNDGPAHPROARREDIEQYLTQNIQYPMSPA 976

RESULT 3

AGOL ARATH STANDARD; PRT; 997 AA.  
AC Q9SJL3;  
DT 01-DEC-2000, integrated into UniProtKB/Swiss-Prot.  
DT 01-MAY-2006, entry version 1.  
DT 07-MAR-2006, entry version 29.  
DE Argonaute-like protein At2g27880.  
GN OrderedLocusNames=At2g27880; ORFNames=TFE2.20;  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxId=3702;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083487; PubMed=10617197; DOI=10.1038/45471;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Rongning C.M., Koo H.L.,  
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Unayam L.,  
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
RA Nierman W.C., White O., Bisen J.A., Salzberg S.L., Fraser C.M.,  
RA Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana";  
RL Nature 402:761-768 (1999).  
CC -!- SIMILARITY: Belongs to the argonaute family.  
CC -!- SIMILARITY: Contains 1 PAZ domain.  
CC -!- SIMILARITY: Contains 1 Piwi domain.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License

DR EMBL; AC006929; AAD21514.1; -; Genomic\_DNA.  
DR PIR; A84678; A84678.  
DR GenomeReviews; CT485783\_GR; AT2G27880.  
DR TAIR; At2g27880; -.  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Piwi.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Piwi; 1.  
DR PROSITE; PS50821; PAZ; 1.  
DR PROSITE; PS50822; PIWI; 1.  
FT CHAIN 1 997 Argonaute-like protein At2g27880.  
FT /FTId=PRO\_0000194070.  
FT PAZ.  
FT DOMAIN 359 471  
FT DOMAIN 638 958 Piwi.  
FT SEQUENCE 997 AA; 111135 MW; DA93A1446C4222P31 CRC64;  
Query Match 11.3%; Score 604.5; DB 1; Length 997;  
Best Local Similarity 23.8%; Pred. No. 31e-25;  
Matches 242; Conservative 164; Mismatches 353; Indels 259; Gaps 41;  
QY 38 KKVLLLVNMFKFSKTYDREYYEYVMTKEVLNRKPKFPKPKTEIPIPRAKLFWOHL 97  
Db 158 KKVMVRAN--HFLVQVADRDLVHYDVSINPEVSK----- 190  
QY 98 RHEKQTDFILEDYVDEKDTVYVCRLLNTVTSKMLVSEKVKVKDSE--KKDEKDLKK 154  
Db 191 -----TVARNVNMKLLVKN---YKDSHLGKSPAYD-GRK 220  
QY 155 ILVTM--ILTYRKPHLNFSENPEKDEEANRSYKF-LKNV-----MTOKVRYA 200  
Db 221 SLVTAGLPFDSKEFVNVLAERKADSGSKORFVKAVKNTSTDLVQLQOFLDRKOREA 280  
QY 201 PNVNEBIKQVAKNFYVYDNNLSILRVPEFHPNRFQSL-----EVAPRIEAWFGI 251  
Db 281 PY--DTIQVLDVLRDKPSNDYVSVGRSP-----FHTSLGKDARDGRGELGDGIEYWRGY 333  
QY 252 YGIKELFDGEPVINFVVDKLPYNAPKMSLLDYLLIIVDPQSCNDVDRKDKTKLMAGK 311  
Db 334 FOSLR-LTQWGLSLNIDVSARSFY-----EPIVVTDFISKFLNIRDLL--N 375  
QY 312 MTIROAARPRIOLLENLKLCAEVWVDNEMSLRHLTLFELDLCEENSLVVKVTKGSD-- 369  
Db 376 RPLRDSRLKVKVKVLRLLKVLH-WN-----CTKSA---KISGISLP 415  
QY 370 ----RGRNAKVDYTLFKIYEENKCF-IEFPHLPLVKVKSGAKEYAVPMHEHVEHPQR 424  
Db 416 IRELRLTLEDKSEKTVVQYFAEKYVYVQALPAIQGSDTRPVVLPMLCQIDE-QGR 474  
QY 425 YKNRIDLVMQDKELKATRKPHDYKENTLKMELDFSEELNPFVERFGLCSKLMIECP 484  
Db 475 YTKRLNEKQVTTALLKATCQRPDP-RENSIKNLVVKNNYNDLDS--KEFGMSVTTQLASIE 531  
QY 485 GKVLKEPMLVNSVNEQIKMTPVIRG----FOEKQALNVVPEKELCCAVFVNVNATAGNCPLE 540  
Db 532 ARVLPPPMKLYHDSGKEKVNPRLGOWNMDKQWNGAKVTSWTCVSFSTRIDRGLP--- 588  
QY 541 ENDVVKFYTELIGGCKFRGIRIGANENRGAQSIIMYDATKNEYAFYKNCITLNTIGRFEIA 600  
Db 589 ----QEFCKQLIGMVCVSKGM----EFKQPAIPFISCPPEH-----IEEA 625  
QY 601 ATEAKMFERLPDKEOKVLMPIISKRQLNAVGVFKHYCDHTIGVANQHITSETVTYKALA 660  
Db 626 LLD---IHKRAPGLQ----LLVILPDVTGYSYKIKRICETELGIVSQCCQVQVKNL-- 676  
QY 661 SLRHEKSGKRIFYQIAKLNKAKLGGINQELDSWETAIEISPEEKERKMTPL-----TMVY 715  
Db 677 -----NKQYMNVAKLNKVTGGRNTVLN-----DAIRNRIPIIDTRPTIIM 718  
QY 716 GIDVTHPTSYSGIDYSTAAVVASIN-PGGTYIRNMIVTQ---EE-----CRPGRVAV 763  
Db 719 GADVTHPTQGEDSSPSIAAVVASMDWPEINKYKGLVSAQAHRREEIIQDLYKLVQDPQRL 778  
QY 764 AHGRETDILEAKFVKLLREFAEENNDRAPAHIVVTRDGVSDSEMLRVSHDELRSLSKSEV 823



Qy	153	KKILYTMILTYRKXKPHLNFSRENPEKDEBANRSYKFLKNVMTOKVYAPFVNEEIKVQFA	212
Db	158	VSIKWLAIWSRWMLHEALVSGGPIVPLESVQALDVMARHLAS--MEYTVV-----	175
Qy	213	KNFVVDNNSILVPBSFHDPNRFEQSLSEVAPRIEAMFGIVIGIKELPDCEPVLNPAIVDK	272
Db	176	-----GRSFFSPPEGYHP-----LGGGEVWFGFHQSVRPAM-WKMWLNLIIVSAT	220
Qy	273	LFYNAPKMSLLDYLLIIVDPQSCNDVVRKDLTKMLMAGKWTIRQAAARPIRQLLENLKLK	332
Db	221	AFYKA--QPVI8FMCVELDIRNIDEPK-----PLTDSQRVFTKEIKGLKVE	266
Qy	333	---CAEVDN---EMSLRTERHLIFDLICEENSUVYKVTGSKDRGNNAKYDTTLFKYI	385
Db	267	VTHCGOMKRYKVCNVRTRPASHQTFPFOLESGQTV-----ECTVAQYF	310
Qy	386	BENKIF-IEFFPHLPVVKVSGAKYAVPMPEHLEHVIEKHPORYKNRIDLVMDQKFLKRATR	444
Db	311	KQKNLQLKYPHLPCLQVQEQEKHYFLFLEVCI--VAGORCICKLTDNOTSMIKATARS	369
Qy	445	PHDYKENTLKMELKELDFSSEELNFERFGLCSKLQMECPGKVLKEPMLVNSVEIQMT	504
Db	370	APDROEBISRLMKNASYNLDP--YIQEFGIKVKDMDTEVTRVLPAPILQYCGRRAJAT	427
Qy	505	P-----VIRGFOEKQL-NVVPEKELCCAVFVNNETAGNCPLEENDVVKYETELI	552
Db	428	PNQGVMDRG---KOFYNGIEIKWAIACFAPQK---OCREE--VLKNFTDQLRKISKD	478
Qy	553	GGCKPRGIRIGANENRGAQSIIMYDATKNEYAFYKNCNTLTGTGIRPEIAATEAKNMFERLP	612
Db	479	AGMPFOGQPCFKYAQADSV-----EPWFRHLK	507
Qy	613	DKEQKVLMPILISRKQLNAYGVFKYCHDHTIGVANQHITSETVTTKALASLRHEKSGKRIF	672
Db	508	NTYSGQLIIVILPKGTPVYAEVKRVGDTLLGMATQCVQKVNVKI-----SPQTL	558
Qy	673	YQIALKINAKLGGINOELDWSIAISLPEEKERRKTMPLTMVYGVGDVTHPTSYSGIDYSI	732
Db	559	SNLCLKINVKLGGINNLI-----VPHQSAVFQOPV-IFLGADVTHPPAGDGKKPSI	609
Qy	733	AAVVASINPGGTIYRNMTVTOBECRPGERAVAHGERTDILS--AKFVK-LLREFAENND	789
Db	610	TAVVGSMDAHP8RYCATVRVQ-----RPROEIIEDLSVMWRELLIQYKSTR	656
Qy	790	NRAPAHIVVYRDGVSDSEMLRVSHDRLSLKSEVKQFMSERDGEDPEPKYTFIVIQRHN	849
Db	657	FK-PTRIIFYRDGVPEGQLPQILHYELLAIRDACIKL-----EKDYQPGITVIVQKRHH	710
Qy	850	TELLARMEKDPVNVKDLTPAETDVAAVAQWEEDMKESKETGIWNPSGGTIVDKLIVS	909
Db	711	TRLF-CADKNERIGKSGNIPA-----GTTVDNTNTH	740
Qy	910	KYKEDFFLASHHGVLGTSRPHGYTMVYDDKGMODEVVKYKMTYGLAFLSARCRRKPSLVP	969
Db	741	PFEDFPYLCSHAGIOGTSRPSHYVWLDNRFTABELQLITLQOLCHTVYCTRVSIPAP	800
Qy	970	VHYAHLSCSEKAKELRYTYKEHYIGD	994
Db	801	AYYARLVAFRAR-YHLVDKEHDSGE	824

RESULT 5  
Q4RKH3 TETNG  
ID Q4RKH3 TETNG PRELIMINARY; PRT; 863 AA.  
AC Q4RKH3;  
DT 19-JUN-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome 21 SCAP15029, whole genome shotgun sequence. (Fragment).  
GN ORFNames=GSTENG003295501;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
NC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed15496914; DOI=10.1038/nature03025;  
 RA Jallion O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Barnot A.,  
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dosat C., Segurens B.,  
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
 RA Bionnet C., Skalli Z., Catcolico L., Poulain J., De Berardinis V.,  
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McSwan P., Bosak S.,  
 RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.:  
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 the early vertebrate proto-karyotype."  
 RL Nature 431:946-957(2004).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; CAAB01015029; CAG11109.1; -; Genomic DNA.

-----  
 Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 Distributed under the Creative Commons Attribution-NoDerivs License  
 -----  
 CC EMBL; CAAB01015029; CAG11109.1; -; Genomic\_DNA.  
 CC SMR; QARKH3; 206-373.  
 CC InterPro; IPR003100; PAZ.  
 CC InterPro; IPR003165; Pwi.  
 CC Pfam; PF02170; PAZ; 1.  
 CC Pfam; PF02171; Pwi; 1.  
 CC PROSITE; PS50821; PAZ; 1.  
 CC PROSITE; PS50822; Pwi; 1.  
 FT NON\_TER 1  
 FT NON\_TER 863  
 SQ SEQUENCE 863 AA; 98054 MW; 8DA4C8A0741D8E183 CRC64;

Query Match 10.7%; Score 573; DB 2; Length 863;  
Best Local Similarity 23.2%; Pred. NO. 1.5e-23;  
Matches 240; Conservative 17%; Mismatches 349; Indels 268;  
Gaps 46

```

19 QY PEMKWLARPTKCDGKFYEKKVLLVNNFKFSSKIYDREYIEYVKMTKEVLNRKPKGPF 78
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1   PVPEVVFPPSRPFGTNGRTIKLQANFEM--EIPKLEVVHYDIDI-----KEK-C 50
    QY 79 PKKTEIPIPDRAKLFWQHLRHEKKQTDFILEYVFDEKDTVYVSVCLRLNTVTSKMLVSEKV 138
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
51 PRVRNREIVEHM-----VQHFKTQ-----IFGDRKPVVD----- 79
    QY 139 VKKSEKKDEKLEKILLYTM--ILTYRKPHLNSRENPEKDEANRSYKFLKNVMTQK 196
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
80 -----GRKNLYTAMPUPIGRDKVELEVTIPGEGKDRSKFKVSIKWSVCSLQA 126
    QY 197 VRYA-----PFVNEBIKVQAKNFYVDNNSILR---VPSEFHDNPNRFEOSLEVPARI 245
    Db 127 LHEALSGELPSVPP--ETVQ---ALDVVWRHLPSMRYTPVGRSFPTPEGCANPLGGPR- 180
    QY 246 EAWFGIYIGIKELPDGEPVLNFAIVDKLFYNAPKMSLLDYLLI VDPSCNDVDVRLDKLT 305
    Db 181 EWVFGFHQSVRPSL-WKMWMLNDVSATAFYKA--QPVIETFCVLEVDKSIIBEQ----- 230
    QY 306 KLMAGKMTIROAAPRIRQLLENLKLK---CAEVWDN-----EMSLRTERHLTFDLCEBN 358
    Db 231 -----QKPLTDSQRVKFTKEIKGLKVEITHCGOMKRKYRVGNVTRRPPASHOTFP-PLQOBN 284

```





DT 21-FEB-2006, entry version 11.  
DE Putative leaf development and shoot apical meristem regulating  
DE protein.  
GN Name=OSJNB0070009.6;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
RA Overton II L.L., Tsirlin T., Kim M.M., Bera J.J., Jin S.S.,  
RA Padrosh D.W., Tallon L.J., Koo H., Zismann V., Heisao J., Blunt S.,  
RA Vanaken S.S., Riedmuller S.B., Utterback T.I., Feldblyum T.V.,  
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
RA White O., Salzberg S.L., Fraser C.M.;  
RT "Oryza sativa chromosome 3 BAC OSJNB0070009 genomic sequence.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RA Buell R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AC087412; AAP68386.1; -; Genomic DNA.  
DR Gramene; Q7Y001; -;  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Pfam.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Pfam; 1.  
DR PROSITE; PS0821; PAZ; 1.  
DR PROSITE; PS0822; Pfam; 1.  
SQ SEQUENCE 1055 AA; 115887 MW; 1817D71A0B71B3DD CRC64;

Query Match 10.6%; Score 568.5; DB 2; Length 1055;  
Best Local Similarity 22.4%; Pred. No. 3.6e-23;  
Matches 226; Conservative 169; Mismatches 345; Indels 271; Gaps 39;

QY 49 FSKYIDREYVEYKMTKEVLRNPKGPKFPKTEIPDRAKFWHLRHEKKQTDFFL 108  
DB 213 FLVQVADKDIYHVDVITPESTYVRNRSIINKL-----VALHKQFLDGLR 258  
QY 109 EYVDEKDTVTSVCRNLNTVTSKMLVSEKVKVDEKDE-----KLEKKILYTMILTY 163  
DB 259 P--VYDGRKSIYTAGPLPKTKDFVVKHINPLRGNQRESEYKVTIKQASKTDLSL---- 312  
QY 164 RKKEHLNFRGRENPEKDEBANRSYKPLKNVMTQKRVYAPVFNBEIKVQPAKNFVYDNNSTL 223  
DB 313 -KQFLVGRQRELPQDTIQA-----LDIALRECP-----TSVNFCTDR--YV 350  
QY 224 RVPESFHPDNRPEQSL-----EVAPRIEAFNGIY-----IGIKEL 258  
DB 351 SISRSE-----FSQFGHGEIGSGTECHRGYYSRLPTQGLSLNIGMDLPQNISATAF 405  
QY 259 FQGEVPLNPAIVDKLPYNAPKMSLLDYLILLIYVDPSQNDVDRKDLKTKLMAGMTIROQA 318  
DB 406 YKAQPMDFPAV-----QYLN-----RDVSRRL-----SDQD 432  
QY 319 RPRIRQLLENLAKCAEVWDNEMSRITERHLTFLDICEENSLVYKVTGKSDRGRNAKYD 378  
DB 433 RIKLKALKGVQI-VATHW-----KEKSIYKLTGTPSPAMNLMFDP 473  
QY 379 TT-----LFKIYEENKFIETPHPLPLVKSGAKEYAVPMHELVHEKPPQYKRI 429  
DB 474 LQGNRISVVQYFKQYNSLKHVNPCL--QAGSDSRPKYLPMEVCSILE-QQYSKK 529  
QY 430 DLVMQDKFLKRAIRKPHDYKENTLQMLKELDFSSSELNFRFGLCKLQMTCEGKVLK 489  
DB 530 NEHQVTNLRLMTCERPAQRESSIIIEIVNTNSYGNDDC--AKFGIKVANQLAVVDARVLP 587

## RESULT 9

QY QP241 ANOGA  
ID Q7P241 ANOGA PRELIMINARY; PRT; 891 AA.  
AC Q7P241;  
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.  
DT 07-DEC-2004, sequence version 2.  
DT 07-FEB-2006, entry version 8.  
DE ENSANGP0000008896 (Fragment).  
GN ORFNames=ENSANGG0000008700;  
OS Anopheles gambiae str. PEST.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae;  
OC Anophelinae; Anopheles.  
OX NCBI\_TaxID=180454;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PEST;  
RG The Anopheles gambiae Sequence Committee;  
RT "Anopheles gambiae re-annotation.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PEST;  
RG The Anopheles gambiae Sequence Committee;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AA0801008986; EAA00062.2; -; Genomic DNA.  
DR SMR; Q7P241; 239-406.  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Pfam.



DR	Pfam:	PF02170; PAZ; 1.
DR	Pfam:	PF02171; Pwi; 1.
DR	PROSITE:	PS50821; PAZ; 1.
DR	PROSITE:	PS50822; PIWI; 1.
FT	NON TER.	1
SQ	SEQUENCE	891 AA; 100211 MW; D2562B84B430F46B CRC64;
Query Match      10.6%; Score 567.5; DB 2; Length 891;		
Best Local Similarity    22.6%; Pred. No. 3.2e-23;		
Matches 206; Conservative 161; Mismatches 346; Indels 199; Gaps 29		
QY	146	KDKDLKKILTYMLTYRKRP-----HINFSRENPEKD 179
DB	83	KCPKRVNREIETWVHAYSKMGALKPVDGRNNLYTRDLLPIGNDRVELEVTLPGEGKD 142
QY	180	BEANRSYFKLVN-----MTOKVRYPAPFYNEIKVGFAGNFVYD--NNSILRPV 226
DB	143	RVPRTVIKWAVQSULFNLEALEGHTROIPIYDAIALDVMMRHLSMYTTPVGRRFFSSP 202
QY	227	ESFHDPNRFOSLEVAPRIEAWFGIYGIGELFDGEVLNFAIVDKLFYNAPMNLSDLY 286
DB	203	DGYHP-----LGGCREVMFGPHQSVRP-SQWKMLNIDVSATAFYA-QPVIERM 251
QY	287	LLIVDPQSCNDVRKDLTKLMAGKWTIRQAARPIRQLLENLKLCABVDNENMSRLTE 346
DB	252	CEVLDIRDINEQ-RKPL-----TDSQRVKPTKEIKGLKTEITH----- 288
QY	347	RHLFTLDICEENSIVKYVTGKSDRG-----RNAKDYDTLFLKIY-BENKXFIRBP 395
DB	289	-----CGTWRRKRYCNVTRRPAQMSPFLQENGQTVCETVAKYFLDKYMKGLRYP 340
QY	396	HLPVKKVSGAKEYAVPMHEHLEHVHKPORIKNRIDLVMQDKPKCATRKPDHYKENTLKM 455
DB	341	HLPCLOVQEHHKHTYLPLEVCNI-VAGQCINKLTDMQTSMTWKATARSAPDREINNL 399
QY	456	LKELDFFSEELNFERFGLCSKLQMIIEGCKVKLKEPML-----VNSVNEIQIMTVIRGF 510
DB	400	VRRADFNDA--YVQEFGLTISNNMEVRGRVLPPLPKLYGGRVSMGSQL-----LSP 452
QY	511	QEQQLNVVPEKELCCAVFVNNETAGNPCLEENDVVKFYTELIGCGCKFGIRIGANENRGA 570
DB	453	QNKVSLALPNQG-----VMDMRGK-----QFFT-----GVEIRWALACPAPQ-- 490
QY	571	QSIWDYDATKNBYAFYKNCNTLNTG---IG-----RPEIAATEAKNNFERLPQKEQVLMFI 622
DB	491	RTVEDALRNFTQOLQKISNDAGMPIIGQPCFCYATGPDQVBPFRVYLKSTFSLQLVV 550
QY	623	IISKRQLNAYGVKHCHDTTGVANQHITSETVTKALASLRHKGSKGIIFYOIALKINAK 682
DB	551	VVLPGKTPTYAEVRXRGVDTVLGMATCQVQANKVNT-----SPQTLSNLCLKINVK 601
QY	683	LGGNQELDWSEIAEISPEEKERRKTMPLTMVYGDIVTHPTSYSIGIDYSIAAVVASINPG 742
DB	602	LGGINSIL-----VPSIRPKVFDEP-----VIFLGADVTHFPAGDNKKPFSIAAVGSGMDAH 652
QY	743	GTIYRNMIVTQECRPGERAAVAGRERTDILEAKFVKLLRFEANNDRNRAPIHVVYRDG 802
DB	653	PSRYAATVRVOQ-----HRQEIQLSSMWREILLIMFYKSTGGFKPHRIILIYRDG 702
QY	803	VSDSEMURVSHDELRLSKSEVKQFMERSGEDPEPKYTFVIQKRHNTRLRLRMKKXPV 862
DB	703	VSEGQFPHVLQHELTAIREACIKLEA----DYKPGITFIVVQKRHHRTLFCADKKE--- 754
QY	863	VNKDLTPAETDAVAAVQKWEEDMKESKETGIIVNPSSCTTVDKLI VSKYKDFPFLASHHG 922
DB	755	-----QSKSG--NIPAGTTVDVGTTHPTETFDFYLCSHQG 787
QY	923	VLGTSRPGHYYMVDDKGMQSDVVKYMTYGLAFLSARCCKPTSLPVPVHYAHLSCAKE 982
DB	788	IQGTSRSHYHVLWDDNHFESEDELQCLTYQLCHTVYTRCTRVSII PAPAYHAHLVAFRAR- 846
QY	983	LYRTYKHEYIGD 994 : : :

847 YHLVEKHSCE 858

RESULT 10  
Q6DCX2\_XENLA PRELIMINARY; PRT; 862 AA.  
Q6DCX2\_  
16-AUG-2004, integrated into UniProtKB/TrEMBL.  
16-AUG-2004, sequence version 1.  
07-FEB-2006, entry version 12.  
Eif2c1-prov protein.  
Name=eif2c1-prov;  
Xenopus laevis (African clawed frog)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
OC Xenopodinae; Xenopus; Xenopus.  
NCBI\_TaxID=8355;  
[1]  
NP\_NUCLEOTIDE SEQUENCE.  
RP TISSUE=Embryo;  
RC MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RX Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT initiative.";  
RL Dev. Dyn. 225:384-391 (2002).  
[2]  
NP\_NUCLEOTIDE SEQUENCE.  
RP TISSUE=Embryo;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fanev J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallus D.E.,  
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

Query Match 10.6%; Score 565.5; DB 2; Length 862;  
Best Local Similarity 24.2%; Pred. No. 4e-23;  
Matches 234; Conservative 151; Mismatches 371; Indels 209

BEST LOCAL SIMILARITY	24.2%;	Pred. NO. 4e-23;
Matches 234;	Conservative 151;	Mismatches 371;
Indels 209;	Gaps 40;	



```
QY 49 FSSKIYDREYVEYEVQMTKEVNLNRKPKPPKTEIPIPDRAKLFWQHLRHEKKQDTFTL 108
Db 47 FEMDIPKILYHYDIDI-----KPEK-CPRVRNREIYEHM-----VQHFAQ---IP 89
QY 109 EDY--VFDEKQTVYVSVCLNTVTSKMLVSEKVKVQKDEKDEKLEKKILYTMILTYRKK 166
Db 90 GDRKPVDFGRKNLYTAMPLPIARDKQV--ELEVTLPGEGKDR--IPKVAIKMACVSLQA 145
QY 167 FHLNFSRENPEKDEENRSYKFLKNVMTQ--KVRVAPFVNEBIKQVQFKNFYDNNLSILR 224
Db 146 LHDALSGRLPNVPFE---TVQALDVVMRHLPSMRYTPV-----GRSFFT 186
QY 225 VPESFHPDNRPFQSLSEVAPRIEAWFGIYIGIKELFDEGPVLNFAIVDKLFYNAPKMSLLD 284
Db 187 ASEGCANP-----LGGREVFWFGFHQSVPRL--WKWMLNIDSATFYKA--QPVIE 235
QY 285 YLLILVDPQSCNDVVRKDLTKMLMAGKMTIROQAARPIRQLLENKLGK---CAEVDN-- 339
Db 236 FMCEVLDFKSIIEQ-----QKPLTDSQRVKFTKEIKGLKVEITHCGQMKRKYR 283
QY 340 --EMSLTERHLTFLDLCEENSLVYKVTGKSDRGNAKKYDTTLPKIYEENKKFIEFPHL 397
Db 284 VCNVTRRAPASHQTF--PLOSQSGTVECT-----VAQY-----FK--DRHKLVLRYPHL 328
QY 398 PLVKVKSAGKAYAVPMHELVHEKQRYVKNRIDVMQDKFLKRAT--RKHDPDYKENTLKM 456
Db 329 PCLQVQGEQKHYLPLEVCNIVAGQRCIKLTD--NOTSTMIRATARSAPDQEIISKUM 386
QY 457 KELDFSSEELNPFVERGLCSKLOMIECPGKVLKEPMLVNSNEQIKMTVPV--IRGFQEK 514
Db 387 RSASENTDP--FVREFGIMVKDDMTDVTGRVLQPPSILYGRSKAIAITPVQGVMDNRNQ 444
QY 515 LNV-VPEKELCAVFNWETAGNPLEENDVVKFTELI-----GGCKFRGIRIGANEN 567
Db 445 FHTGIEIKVWAIACFAPO---RQTEVH--LKTFEQLRKISRDAGMPIQGPCCKYA 498
QY 568 RGAQSI--MYDATKNEYAFYKNCNTLNTGIGRFEIAATAKNMFERLPDKEQKVLFAIIS 625
Db 499 QGADSVPMFRHLKNTY-----TGL-----QLVVIL 525
QY 626 KQLNAYGVFVGHYCDHTIGVANQHITSETVTTKALASLRHEKSGKRIFYOIALKINAKLGG 685
Db 526 PGKTPVYAEVKRVGDTVLMGATQCQVMKNV-----QRTTPQTLNLCIKINVKLGG 576
QY 686 INQELDWSIAISISPEEKERRKTMPLTWYVGIDVTHPTSYSGIDYSIAAVASINPGGTI 745
Db 577 VNNIL-----LPQGRPPVFQPV-IFLGADVTHPPAGDGKPKSIAAVGMSMAHPNR 627
QY 746 YRNMIVTQECRPGERAVAGHRTDILEAKFVKLLRFAENNDNRAPAHIVVYRDGVSD 805
Db 628 YCATVRVQ-----HREIIQDLSAMVRELLIQYKSTRPK-PTRIIFYRDGVSE 676
QY 806 SMLRVSHDELRSLSKVKQFMSRDEGDEPEPKYTFIVIQKRNHLLRRMEKOKPVNKK 865
Db 677 GQFQVQLVHLLAIREACIKL-----EKDYQPGITFIVVQKRHHTLRF-CTDRNERVGKS 730
QY 866 DLTRAETDVAAVAVQWEDMKESKETGIVNFSSTGTVDKLIVSKYKDFPFLASHHVLG 925
Db 731 GNIPA-----GTTVDTKITHSEDFDYLCSHAGIQG 761
QY 926 TSPRGHYTVMYDDKGMQDEVYKMTYGLAFLSARCKRKPISLPVPVHYAHLSCAKELYR 985
Db 762 TSPRSHYVLDNDRFSSDELQILYQLCHTVYVTRTSVSIIPAYAYAHVAFRAR--YHL 820
QY 986 TYKEH 990
Db 821 VDKEH 825
RESULT 11
12C2 RABIT
ID 12C2 RABIT
AC 077503; PRT; 840 AA.
```

```
DT 01-DEC-2000, integrated into UniProtKB/Swiss-Prot.
DT 14-NOV-2003, sequence version 2.
DT 07-FEB-2006, entry version 30.
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
GN (Argonaute-2) (Fragment).
OS Name=eIF2C2; Synonyms=AGO2;
OC Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha; Leporidae;
OC Oryctolagus.
OC Oryctolagus.
NCBI_TaxID=9986;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND PARTIAL PROTEIN SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=98267198; PubMed=9602122; DOI=10.1016/S0378-1119(98)00107-3;
RA Zou C., Zhang Z., Wu S., Osterman J.C.;
RT "Molecular cloning and characterization of a rabbit eIF2C protein.";
RL Gene 211:187-194(1998).
CC -!- FUNCTION: Provides endonuclease activity to RNA-induced silencing
CC complexes (RISC). Cleaves siRNA/mRNA heteroduplexes bound to RISC.
CC Essential for embryonic development as well as RNA-mediated gene
CC silencing (RNAi) (By similarity).
CC -!- SUBUNIT: Interacts with DICER1 through its Piwi domain (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic bodies (By similarity).
CC -!- DOMAIN: Piwi domain is thought to provide RNA cleavage activity
CC with a mechanism similar to RNase H (By similarity).
CC -!- SIMILARITY: Belongs to the argonaute family.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC -----
CC EMBL; AF005355; AAC24323.1; ALT_INIT; mRNA.
DR PIR; PC6505; JC6569.
DR SMR; O77503; 203-370.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Direct protein sequencing; Initiation factor; Protein biosynthesis.
FT Eukaryotic translation initiation factor
FT 2C 2.
FT /FTid=PRO_0000194059.
FT PAZ.
FT DOMAIN 216 329
FT DOMAIN 498 799
FT NON_TER 1
FT SEQUENCE 840 AA; 95306 MW; 1E703F9E31391F29 CRC64;
SQ
Query Match 10.6%; Score 565; DB 1; Length 840;
Best Local Similarity 23.2%; Pred. NO. 4.1e-23;
Matches 231; Conservative 157; Mismatches 357; Indels 250; Gaps 39;
QY 38 KVVLLVNVWFSSKIYDREYVEYEVQMTKEVNLNRKPKPPKTEIPIPDRAKLFWQHL 97
Db 17 RTIKLQANFFEMDIPKIDYHYELDKPEK-----CPRVRNREIYEHM-----V 60
QY 98 RHEKKQDTFTILEDY--VFDEKQTVYVSVCLNTVTSKMLVSEKVKVQKDEKDEKLEKKI 155
Db 61 QHFAQ---IFGDRKPVDFGRKNLYTAMPLPIGREKV---ELEVTLPGEGKDR--IFKVS 112
QY 156 LYTWILTYRKKFHLNFSRENPEKDEENRSYKFLKNVMTQ--KVRVAPFVNEBIKQVQFNAK 213
Db 113 IKWVSCVSLQALHDALSGLPSVPPE---TIQALDVVMRHLPSMRYTPV----- 158
QY 214 NFVYDNNISILRVPSFHPDNRPFQSLSEVAPRIEAWFGIYIGIKELFDEGPVLNFAIVDKL 273
Db 159 -----GRSFFTASEGCSNP-----LGGREVFWFGFHQSVPRL--WKWMLNIDSATA 204
QY 274 FYNAPKMSLLDYLLIIVDPQSCNDVVRKDLTKMLMAGKMTIROQAARPIRQLLENKLGK- 332
```





```

Db 254 TDSORVKTKEIKGLKVEITHCGQMKRYKRVNTRTPASHQTF-PLQESGQTVECT-- 310
Qy 368 SDRGRNKKYDVTTLFKIYENKKEIFBPHPLPVKVKSGAKAYAVPMHLEHVEHPQRYKN 427
Db 311 -----VAQY-----FK--DRHKLVLRYPHLPCLVQGOEQKHLYLPLEVCNIVAGQRCIKK 358
Qy 428 RIDLVMDQKFLKRAT-RKPHDYKENTLKLKELDFSSSEELNFVERFGLCSKLQWIECPGK 486
Db 359 LTD--NOTSTWIRATARSADPRQBEISKMRASFNTPD--FVREFGIMVMDQDMDTWTGR 414
Qy 487 VLKEPMLV-----NSVNEQIKMTPV--IRGFQBEKQLMV-VPEKELCCAVFVNNETA 534
Db 415 VLQPPSILYGRVWEEBNAFLNKAIAITPVQGVMDMRNKQFHTGTEIKVWALACFAPQ-- 471
Qy 535 GNPCLENDVVKFTTELI-----GGCKFRGIRIGANENRGAQSI--MYDATKNEYAFYK 586
Db 472 -ROCTEVH--LKTFTQRLKISRISDAGMPIQGPCFKYACQADSVEPMFRLKNTY---- 524
Qy 587 NCTLNTGIGRFEIATAEAKMFERLPDKEQVLMFIIISKQLNAYGVKHYCDHTTIGVA 646
Db 525 -----TGL-----QVUVVILPGKTPVYAEVRKRVGDTVLGMA 555
Qy 647 NQHTSETVTKALASLREHSGSKRIFYQIALKINAKLGINOELDWSEIAEISPEEKERR 706
Db 556 TQCVQMKV-----QRTTPTLSNLCKINVLGGVNNIL-----LPQGRPPV 598
Qy 707 KTWPLTWYGVIDWTHPTSYSGIDYSIAAVASINPGGTIYRNIMVITQEBRCRGERAVAHG 766
Db 599 FOQPVV-IFLGADVTHPPAGDGKPKSPAAVVGSDMAHPNRYCATVRVQQ-----HR 647
Qy 767 RERTDILEAKFVKLLREFAEANNDRAPAHIVVVRDGVSDSEMLRVSHDELRLSKSEVKQF 826
Db 648 QEIQLDSAMVRELLIQFYKSTRFK-PTRIIFYRDGVSEGOFOQVLHHELLAIREACIKL 706
Qy 827 MSERDGEDPDKYTFIVIOQRHNTLRLRRMEKDKPVVVKDLPATETDVAVAAVKQWEDM 886
Db 707 -----EKDYQGITFIWQGRHTRLE-CTDRNVRGKSGNIPA----- 744
Qy 887 KESKFTGIVNPPSGTTVDKLVISKYKPDFFLASHGVGLTSRPHGYTVMYDDKMGSDQEV 946
Db 745 -----GTTVDTKITHPSBDFLYLCSHAGIQGTSRPSHYHVLWDDNRFSDEL 791
Qy 947 YKMTYGLAFLSARCRPISIPVHVYHLSCEKAKELYRYTYKEH 990
Db 792 QILTYQLCHTYVRCRVSVPAPAYYAHVAFR-YYHLVDKXH 834

RESULT 14
ID 12C2 RAT STANDARD; PRT; 860 AA.
AC Q9QZB1;
DT 01-DEC-2000, integrated into UniProtKB/Swiss-Prot.
DT 14-NOV-2003, sequence version 2.
DT 07-FEB-2006, entry version 26.
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE (Argonaute-2) (Golgi ER protein 95 kDa) (GERp95).
GN Name:EIF2c2; Synonyms=Ago2;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=Sprague-Dawley; TISSUE=Hepatoma;
RX MEDLINE=99443791; PubMed=10512872;
RA Cikaluk D.E., Tahbaz N., Hendricks L.C., DiMattia G.E., Hansen D.,
RA Pilgrim D., Hoban T.C.;
RT "GERp95, a membrane-associated protein that belongs to a family of
RT proteins involved in stem cell differentiation.";
RL Mol. Biol. Cell 10:3357-3372(1999).
CC -!- FUNCTION: Provides endonuclease activity to RNA-induced silencing
CC complexes (RISC). Cleaves siRNA/mRNA heteroduplexes bound to RISC.

```

```

CC Basal for embryonic development as well as RNA-mediated gene
CC silencing (RNAi) (By similarity).
CC -!- SUBUNIT: Interacts with DICER1 through its Piwi domain (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic bodies (By similarity).
CC -!- DOMAIN: Piwi domain is thought to provide RNA cleavage activity
CC with a mechanism similar to RNase H (By similarity).
CC -!- SIMILARITY: Belongs to the argonaute family.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC -!- CAUTION: Was originally (Ref.1) thought to be membrane-associated.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: AF195534; AAF12800.1; ALT_INIT; mRNA.
CC SRR: Q9QZB1; 223-390.
CC Ensembl: ENSRNOG0000008533; Rattus norvegicus.
CC RGD: 621255; EIF2C2.
CC GO: 0030154; P:cell differentiation; NAS.
CC InterPro: IPR003100; PAZ.
CC InterPro: IPR003165; Piwi.
CC Pfam: PF02170; PAZ; 1.
CC PROSITE: PS08221; PAZ; 1.
CC PROSITE: PS08222; PIWI; 1.
CC Initiation factor; Protein biosynthesis.
KW Eukaryotic translation initiation factor
FT CHAIN 1 860
FT /FTID=PRO_0000194060.
FT DOMAIN 236 349 PAZ.
FT DOMAIN 518 819 Piwi.
FT SEQUENCE 860 AA; 97318 MW; A5B0798C66481C9C CRC64;
SQ
Query Match 10.5%; Score 564; DB 1; Length 860;
Best Local Similarity 23.2%; Pred. No. 4.8e-23;
Matches 231; Conservative 156; Mismatches 358; Indels 250; Gaps 39;
Qy 38 KKVLLLVNMFSSKIYDEYEVKMTKEVLNRKPGKPPFKTEIPIIDRAKLFQHL 97
Db 37 RTIKLQANFEMDIPKIDIVHYELDIKPEK-----CPRVNEIVEHM-----V 80
Qy 98 RHEKQTDFILEDY--VFDEKDTVYVCRINTVTSKMLVSEKVKVKKDSEKKDEKLEKKI 155
Db 81 QHFKTK--IFGDRKEVFDGRKNLYTAMPLIGRDKV--ELEVTLPGEGKDR--IFKVS 132
Qy 156 LYTMILTYYKKPHLNFESRENPEKDEANRSYKELKVVMTQ--KRVYAPFVNEBEIKVQFAK 213
Db 133 IKVSCVSIQALHDAUGRLPSVPFE---TIQALDVVMRHLPSMRYTPV----- 178
Qy 214 NFVYDNNISILRVPESEFHDNRRFEQSLVAPRIEAFGIYIGIKELPDEGPVLNFAIVDKL 273
Db 179 -----GRSFTTASEGCSNP-----LGGREVFEGFHSQSVPSL-WKQMLNIDVSATA 224
Qy 274 FYNAPKMSLLDYLILLVDPQSCNDDVRKDLTKLMAGKMTIROAARPIRQLLENLKLK- 332
Db 225 FYKA--QPVIEFVCEVLDFKSIIEEQ-----QKPLTDSQRVKFTKEIKGLKVEI 270
Qy 333 --CAEWDN----ENSLRTERHLTFLDLCEENSLVYKVTGSKDRGNKAKYDTLTKIYE 386
Db 271 THCGQMKRYKRVNTRTPASHQTF-PLQESGQTVECT-----VAQY-----FK--D 315
Qy 387 ENKGFTEFPPLPLVVKSGAKAYAVPMHLEHVEHPQRYKNRIDLVMQDKFLKRAT-RKP 445
Db 316 RHKLVLRYPHLPCLVQGOEQKHLYLPLEVCNIVAGQRCIKLTD--NQSTMTIRATARS 373
Qy 446 HDYKENTLKLKELDFSSSEELNFVERFGLCSKLQWIECPGKVLKEPMLVNSVNEQIKMT 505
Db 374 PDRQBEISKMRASAFNTPD--YVREFGIMVMDQDMDTWTGRVLQPPSILYGRNKAIAATP 431
Qy 506 V--IRGFQBEKQLNV-----VPEKELCCAVFVNNET-----AGNP 537
Db 432 VQGVMDMRNKQFHTGIEIKVWALACFAPQRC-CTEVHLKSFTEQLKISRDRAGMPIQGGP 490

```

```
QY 538 CLEENDVVKFYTELIGGCKFRGIRIGANENRGAOSI--MYDATKNEYAFYKKNCTLNTGIG 595
Db 491 CF-----CKYA-----QGADSVPEPMFRHLKNTYA-----514
QY 596 RPEIAATEAKNMFERLPDKEQKVLFIITISKQLNAYGFVKYCHDHTIGVANQHITSETV 655
Db 515 -----GLQLVVVILPGKTPYAEVKRGVDTVLGMAQCQVMKNV 553
QY 656 TKALASLRHEKSGKRFYQIOLKINAKLGGINOELDWSIEAISPEEKERRKTMPLTMV 715
Db 554 -----QRTTPQTLNLCKINVLKGVNNIL-----LPQGRPPVFQOPV-IFL 595
QY 716 GIDVTHPTSYSYDVSIAVASINPGGTYIRNMIVTQECRPGERAVAGHERTDILEA 775
Db 596 GADVTHPPAGDGKKSIAAVGSMDAHPRNYCATVRVQQ-----HROEIIQDLAA 645
QY 776 KFKVLLREPAENNDNRAPAHIVVYRDGSDSEMLRVSHDELRLSKSEVKQFMSERDGEDP 835
Db 645 MYRELLIQYKSTRFK-PRTIIFYRDGVSEGQFQVLLHLLAIREACIKLEKEY-----699
QY 836 EPKYTFIVIQKRHNTLLRRMEKDKPVVKNKDLTPAETDVAAVAKQMEEDMKESKETGIV 895
Db 700 QPGITFIVQKRRHTRLF-CTDKNERVKGSGNIPA-----733
QY 896 NPSGGTVDKLVSKYKDPFFFLASHGVLTGRPGHYTVMYDDKMGSDQEVVYKTYGLAF 955
Db 734 -----GTTVDTKITHPTFPDFLYCSHAGIQGTSRPSHYHLVWDNRFRSSDELQILTYQLCH 789
QY 956 LSARCKRPTSLPVVHYAHLSCAKELVRYTKEH 990
Db 790 TVYRCTRVSIIAPAYAHVAFRAR-YHLVDKEH 823

RESULT 15
Q4VAB3_MOUSE PRELIMINARY; PRT; 883 AA.
AC Q4VAB3;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Bif2c2 protein (Fragment).
GN Name=Bif2c2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B5/EGFP transgenic ICR mice; TISSUE=Trophoblast stem cells;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zesberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
```

```
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B5/EGFP transgenic ICR mice; TISSUE=Trophoblast stem cells;
RG NIH MGC Project;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBSJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; BC096465; AAH96465.1; -; mRNA.
DR SMR; QAVAB3; 246-413.
DR MGI; MGI:2446632; Bif2c2.
DR GO; GO:0003743; F:translation initiation factor activity; RCA.
DR GO; GO:0006412; P:protein biosynthesis; RCA.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Pwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Pwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; Pwi; 1.
DR NON_TER 1
SQ SEQUENCE 883 AA; 99333 MW; 6D807B35F5689017 CRC64;
Query Match 10.5%; Score 564; DB 2; Length 883;
Best Local Similarity 23.3%; Pred. No. 5e-23;
Matches 232; Conservative 157; Mismatches 356; Indels 250; Gaps 40;
QY 38 KVVLLVWVFPSSKIYDREYVEYVMTKEVLRKPKPKPKTEIPIPDRAKLFWQHL 97
Db 60 RTIKQANFFEMDIPKIDYHYELDI-----KPGK-CPRRVNREIVEHM-----V 103
QY 98 RHEKKQTDFILEDY--VFDEKDTVYVSVCLNTVTSKMLVSEKVKVKKDKDEKOLEKKI 155
Db 104 QHFKTQ---IFGDRKPFVDGRKNLYTAMPLPIGRDKV---ELEVLPGEKDR--IFKVS 155
QY 156 LYTMILTYRKPFHLNFSRENPEKEBEANRSYKFLKNVMTQ--KVRYAPVNEBEIKVQFAK 213
Db 156 IKWVSCVSLQALHDALSGRLSPVPE---TIQALDVMVHRHLPMSRYTPV-----201
QY 214 NFVYDNNILRVPESEFHPDNPFEQSLVAPRIEAMFGIYIGIKELFDGEPVLNFAVDKL 273
Db 202 -----GRSFFTTASEGCSNP-----LGGGREVWFGHQSVRPSL-WQOMLNIDVSATA 247
QY 274 FYNAPKMSLLDYLLIIVDPQSCNDVDVRKDLTKLMAGKMTIRQAAARPRIRQLLENLKLK- 332
Db 248 FYKA--QPVEFVCEVLDFKSIIEQ-----QKPLTDSQRVKETKIKGLKVEI 293
QY 333 --CAEVWDN-----EMSLRTERHLTFDLCEENSLYKVTGKSDRGNAKKYDTTLPKIYE 386
Db 294 THCGQMKRKYRVCNVTRRPASHOTF-PLQOESGQTVECT-----VAQY----FK--D 338
QY 387 ENKKIEFPHLPVVKVSGAKAYAVPMHELVHEKPYKRNIRIDLVMQDKFLKRAI-RKP 445
Db 339 RHKLVLRYLPHLPCLQVGQEQKHTYLPLEVCNIVAGQRCIKKLTD--NOTSTMIRATPARSA 396
QY 446 HDYKENTLMLKELDFSSSEELNFRVERFGLCSKLOMECPGKVLKPEMLVSNVNEQIKMTP 505
Db 397 PDRQBEISKMLKASAFNTDP--YVREFGIWKDEMTDVTGRVLQPPSILYGGNKAIATP 454
QY 506 V--IRGFOEKQLNV-----VPEKELCCAVFVVVNET-----AGNP 537
Db 455 VQGVWDMENKQPHGTGIEIKVWAIACFAPQRC-CTEVHLKSFTEQLRKISRDAGMPTQGOP 513
QY 538 CLEENDVVKFYTELIGGCKFRGIRIGANENRGAOSI--MYDATKNEYAFYKKNCTLNTGIG 595
Db 514 CF-----CKYA-----QGADSVPEPMFRHLKNTYA-----537
QY 596 RPEIAATEAKNMFERLPDKEQKVLFIITISKQLNAYGFVKYCHDHTIGVANQHITSETV 655
Db 538 -----GLQLVVVILPGKTPYAEVKRGVDTVLGMAQCQVMKNV 576
QY 656 TKALASLRHEKSGKRFYQIOLKINAKLGGINOELDWSIEAISPEEKERRKTMPLTMV 715
Db 577 -----QRTTPQTLNLCKINVLKGVNNIL-----LPQGRPPVFQOPV-IFL 618
```



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 13:34:06 ; Search time 52 Seconds  
(without alignments)  
1716.947 Million cell updates/sec

Title: US-10-645-746-3

Perfect score: 5349

Sequence: 1 MSNPFPEKGFYRSLDPE.....RHEMEHFLQTNVYPMGSPA 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*  
1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/8\_COMB.pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/9\_COMB.pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/10\_COMB.pep.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/11\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	577	10.8	857	2	US-09-949-016-6819
2	577	10.8	894	2	US-09-949-016-10324
3	575.5	10.8	969	2	US-09-533-029-106
4	448.5	8.4	491	2	US-09-270-767-43512
5	427	8.0	1145	3	US-09-866-557A-5
6	391.5	7.3	623	2	US-09-270-767-43950
7	312	5.8	775	2	US-10-043-774B-2
8	306	5.7	861	2	US-09-873-737A-6
9	304	5.7	862	2	US-09-873-737A-4
10	288	5.4	829	2	US-10-104-047-3439
11	266.5	5.0	852	2	US-10-104-047-3646
12	265	5.0	257	2	US-09-248-796A-20922
13	209.5	3.9	844	2	US-10-043-774B-13
14	205	3.8	843	2	US-09-873-737A-2
15	178.5	3.3	245	2	US-09-270-767-44856
16	164	3.1	2954	3	US-09-150-867-1
17	163	3.1	2954	3	US-09-724-584-1
18	163	3.0	1780	2	US-09-949-016-6899
19	163	3.0	1786	2	US-09-949-016-7880
20	162.5	3.0	1244	2	US-09-949-016-11702
21	162	3.0	2662	2	US-09-595-684B-31
22	161.5	3.0	2662	2	US-09-538-092-1252
23	158	3.0	2733	2	US-09-949-016-11433
24	158	3.0	3259	2	US-09-949-016-6507
25	157	2.9	3248	1	US-08-353-700-1
26	157	2.9	3248	5	PCT-US95-16216-1

#### ALIGNMENTS

##### RESULT 1

US-09-949-016-6819  
; Sequence 6819, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 6819

; LENGTH: 857

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-6819

Query Match 10.8%; Score 577; DB 2; Length 857;  
Best Local Similarity 23.0%; Pred. No. 4.7e-43;  
Matches 227; Conservative 171; Mismatches 365; Indels 222; Gaps 37;

QY	38	KKVLLVNMFKSSKIYDREYVEYKMTKEVLRKPGKPPKKTIPIDRAKLFWQHL	97
DB	34	KKPKLLANFEVDIPKIDVHYEVDIKPKD-----CPRVRREVVE-----YMW	77
QY	98	RHEKQTDFILEDY--VFDEKDTVSVCLNTVTSMVSEKV---VKKDSEKKDKDLE	152
DB	78	QHFKPQ---IFGDRKPVYDGNKIY-----TVPALPIGNRVDVFTIPGEGKD-RIFK	127
QY	153	KKILYTMILTYRKCFHLNFSRENPEKDBANRSYKFLKNVMTOKRYAPVFNBEIKVQFA	212
DB	128	VS1KWLAIWSRMLHEALVSGQIPVPLESQALDVAHRHLAS--MRYTPV-----	175
QY	213	KNFVYDNNILRVPSFPHDPNFEQSLVAPRIEAMFGIYIGIKELFDGEPVLNFAIVDK	272
DB	176	-----GRSFFGPPGEGYHP-----LGGGREGVWFGHQSVRPAM-WMOMLINIVSAT	220
QY	273	LFVNAPFMILLDYLLIIVDPQSCNDDVRDKLTKLMAGKMTIRQAARPRIRQLLENLKLK	332
DB	221	AFYKA--QPVIEMCEVILIRNIDEQPK-----PLTDSORVPTKEIKGLKVE	266





Qy	698	ISPEKERKRTKWTPLTMYGVIGDVTHPTSGYDYSIAAUVVASINPGGTIVRNMIVTQEBCKR	757
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Db	210	-VPSTRPWFVNEPV-IFLGADVTHPPAGDNKKPSIAAUVVGSMDAHPSRYAATVRVQQ---	264
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Qy	758	PERAVAHGRERTOLEAKFVKLLREFAENNDRAPAHIVVYRDGVSDSEMLRVSHDELUR	817
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Db	265	-----HRQYIIQELSSWRELLIMFYKSTGGYKPHRIILYRDGVSEGQPPHVLQHELT	317
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Qy	818	SLKSEVQFQMSERGEDDEP--PKYTFIVIQKHNTLLRRMEKQKPVVNKOLTPAETDVA	875
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Db	318	AIREACIKL-----EPEYRPGITFIVVQKHHTRLFCAEKK-----	354
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Qy	876	VAAVKQWEEDMKESKETGIVNPSSGTTVDKILVSKYKDFDFLASHHGVLGTSRPGHYTVM	935
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Db	355	-----QSGKSG--NIPAGTTVDVGITHTEFDFYLCSHQGIQGTSRPSHYVL	400
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Qy	936	YDDKGMGODEVYKMTYGLAFLSARCKRPISLPVPVHYAHLSCAKELVRYTYKHEYIGD	994
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	
Db	401	WDDNHFDSDLOCLTYQLCHTYVTRTSVSPAPAYAHVAFRAR--VHLVEKHDSDGE	458
		::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::	

RESULT 5

US-09-866-557A-5

Sequence 5, Application US/09866557A

Patent No. 7029844

GENERAL INFORMATION:

APPLICANT: Hammond, S.

APPLICANT: Hannon, G.

APPLICANT: Beach, D.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RNA INTERFERENCE

FILE REFERENCE: GNCA-P02-007

CURRENT APPLICATION NUMBER: US/09/866,557A

CURRENT FILING DATE: 2002-03-26

PRIOR APPLICATION NUMBER: 60/189,739

PRIOR FILING DATE: 2000-03-16

PRIOR APPLICATION NUMBER: 60/243,097

PRIOR FILING DATE: 2000-10-24

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn version 3.1

SEQ ID NO 5

LENGTH: 1145

TYPE: PRT

ORGANISM: Drosophila melanogaster

US-09-866-557A-5

```

RESULT 6
US-09-270-767-43950
; Sequence 43950, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43950
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-43950

```

Qy	416	LEVHEKPOYKORIDILVMO--DKFLKRAKTKPHDYKENTLKMUKELDPSSEELN--FVERFG	473
Db	121	CSIEE--GOALNRKDGATOVANMIYAK--PHRRTCESARLXTCNTSTTWDPTISRPG	176
Qy	474	LCSKLOMTECPKVLKEPMLVNSVNEQIKMTPIVRSQBE---KQLNVPEKELCCANFV	529
Db	177	IRIANDFIWSTRVLNPOV---EYHSKRFTPVKNGSWRMQMKLEPKPAHKHCAVLVC	233
Qy	530	-----VNSTAGNPCLEENDVVKFYFTELIGGCKFRGIRIGANENRGAQOIMYDA	577
Db	234	DPRSGRKMYAQLND-FGNLIISQKAVNI---TLDSDDVYRPF---TDDERSLDTIFADL	287
Qy	578	TKNEYAFYKNCITLNTIGRFEIATAEAKMWERLPDKQKVLMFIISKRQLNAGYFVCH	637
Db	288	KRSQHD-----LAIVIIQFRI--SYDTIQ	311
Qy	638	YCDHTIGVANQHITSETVTKALASLRHBKSGKRIEYQIALKINAKLGGINOBLDWSIAE	697
Db	312	KAELOHGILTOCIQFIVER-----KCNQTTIGNILLKINSKLGINHKI-----	356
Qy	698	ISPKEKRRKTMPLTMVVGIDVTHPTSYSGIDYSIAAVASINPGGTYIRNMIVTOBECR	757
Db	357	---KDDPRLPMXGNTWYIGADVTHPSPOREBIPSVVGVAASHDPYGASYNMYRLQ----	409
Qy	758	PGERAVAHGRETDILRAKF---VKLLREFANNDNRAPAHIVYVRDGVSDSEMLRVSHD	814
Db	410	-----RGALSEIDMFITLHLRYKEYR-NAYPDHIIYYRDGVSGQPPKIKNE	459
Qy	815	ELRSLKSEVKOFMSERDGEDPEPKYTFIVIOKRHNTRLRLRMEXDKPVVNKDLTPAETDV	874
Db	460	ELRCIKQAC-----DKVGCKPKICCVIVVKRHTRFPF-----	492
Qy	875	AAVAVKQWEEDMKESKETGIYNPSGGTTVDKLIIVSKYKDFDFLASHHGVLGTSRPHVTV	934
Db	493	-----SGVDTTSNKFNNVDP--GTVVDRTIVHPNEMQPFWMVSHQAIQGTAKPTRXNV	542
Qy	935	MYDDKMGSDQVYKNTYGLAFLSARCKRPISLPVPVVAHLSCAKELXYTYKEHYIG-	993
Db	543	IENTGNLDIDLQQLTYNLCHMFPCNRCSVSPAPAYLAHLVAABGR-VYLTGTNRFDLD	601
Qy	994	--DYAQRTRRHEHFLOTNVKY	1014
Db	602	KCEYAK---RTIVPEFMKKNPMY	621

```

; Patent No. 6723534
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piw1 FAMILY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; FILE REFERENCE: Attorney Docket No. 6723534 180-104/2
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (76)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (303)
; OTHER INFORMATION: Xaa=Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (735)
; OTHER INFORMATION: Xaa=Leu or Ile

```

## US-09-873-737A-6

Query Match 5.7%; Score 306; DB 2; Length 861;  
 Best Local Similarity 23.0%; Pred. No. 2.3e-18;  
 Matches 151; Conservative 88; Mismatches 226; Indels 174; Gaps 26;

QY 417 EVHE-KPORYKNRIDLVMQDKF-LKRATRKPXYKENTLKLKELDFS-----462  
 DB 291 EHKFEQVSKXIGLIVLTKNKYRYVDDIDWDQNPSTFKADGSGVSELYYRQY 350  
 QY 463 SEELNFERGLCSKQMECPKVLKEP-MLVNSVNEQIKMTPVIRGFOEKQLNVVPEK 521  
 DB 351 NQETIDLKQPLVLSQPKRRRGGTLPGPAMLIPELCYLTGLTDKMR-----K 404  
 QY 522 ELCCAVFVNETAGNCP-----LEENDVVK-----FYTELI-- 552  
 DB 405 DLAVHTRLTPEQORQEVGRGLDIYIHKNDVQRELDRDWGLSFDNSLLSFGSRILQTEKHQ 464  
 QY 553 GGCKF---RGIRIGANENRGAOSIMYDATKNEAFYKNCNTLNTGIGRFEIAATEAKNMF 609  
 DB 465 GGTFTDYNQPADWSKETRGAPLISVKPLDNWLLIYTR-----RNYEAANSLLQNLFPK 517  
 QY 610 RLP-----DK-----EOKV-----LMFIIISKRQLNAYGFVHYCDH 641  
 DB 518 VTPAMGMQMRKAIMIEVDRTTEAYLRVLOQKVTDQTQIVVCLLSNRKDKYDAIKKYLCT 577  
 QY 642 TIGVANQHITSETVTKALASLRHEKSGKEIFYOIALKINAKLGGINQELDWSEIAISPE 701  
 DB 578 DCFPTSPQCVVARTLGK-----QOTVMAIATKIALQMNCKMGG---EL-W----- 617  
 QY 702 EKERRKTMF--LTMVYGVIDVTHPTSYSGIDYSIAAVVASINPGGTYIRNMIVTQBECPG 759  
 DB 618 ----RVDIPLKLVMIVGIDYCHDWTAG--RRSIAGFVASINEGTRWFSRCVFQD----- 666  
 QY 760 ERAVAHGRERTDILEAKFVKLLREFAEANDNRPAPAHIVYVRDGVSDSEMLRVSHDELRL 819  
 DB 667 ----RGQELVDGLKVCLOAALRAWNSCNE-YMPSRIIVYRDGVGDGQLKTL-----V 713  
 QY 820 KSEVQKQFMS--ERDGEDPEPKYTFIVIOKRHNTRLLRRMEKDKPVVVKDLTPAETDVA 877  
 DB 714 NYEVQFQFLDCLSKSGVGNPRUTVIVVKRVNTRFF----- 749  
 QY 878 AVKQWEDMKESKETG--IVNPSGGTTVDKLIYSKYKDFDFLASHHGVLGTSRPGHYTM 935  
 DB 750 -----AQSGRLQNLPLPGTVIDVETRPEDWYDFVVSQAVRSGSVSPHYNYI 797  
 QY 936 YDCKGMSQDEVKMTYGLAFLSARCKKPISLPVPVHYAH 974  
 DB 798 YDNSGLKPDHIORLTYKLCCHIYNNWPGVIRVPAPCOYAH 836

## RESULT 9

US-09-873-737A-4  
 ; Sequence 4, Application US/09873737A  
 ; Patent No. 6723534  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Duke University  
 ; APPLICANT: Lin, Haifan  
 ; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE  
 ; FILE REFERENCE: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME  
 ; CURRENT APPLICATION NUMBER: US/09/873,737A  
 ; CURRENT FILING DATE: 2001-06-04  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28764  
 ; PRIOR FILING DATE: 1999-12-03  
 ; PRIOR APPLICATION NUMBER: 60/110,901  
 ; PRIOR FILING DATE: 1998-12-04  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 862  
 ; TYPE: PRT  
 ; ORGANISM: Mus sp.

## FEATURES:

; NAME/KEY: misc\_feature  
 ; LOCATION: (90)-  
 ; OTHER INFORMATION: Xaa=Leu or Ile  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (216)  
 ; OTHER INFORMATION: Xaa=Leu or Ile  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (383)  
 ; OTHER INFORMATION: Xaa=Leu or Ile  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (816)  
 ; OTHER INFORMATION: Xaa=Leu or Ile  
 US-09-873-737A-4

Query Match 5.7%; Score 304; DB 2; Length 862;

Best Local Similarity 22.0%; Pred. No. 3.5e-18;  
 Matches 141; Conservative 101; Mismatches 226; Indels 174; Gaps 24;

QY 414 EHLEVHE-KPORYKNRIDLVMQDKFLKRAATKPH-DYKENTLKLKELDFSSEELNFVER 471  
 DB 289 QOTEHKEQEVSKEGLIGLIVLTKNKYRYVDDIDWDQNPSTFKADGS--EVSFLEY 346  
 QY 472 FGLCSKLQMECPGKVLKEPMLVNSVNEQ-----IKMTPVIRGFOEKQLN-- 516  
 DB 347 YRKQYNQETID-----LKQPLVLSQPKRRRGGTLPGPAMXIPELCYLTGLTDKMRND 401  
 QY 517 -----VVPE---KELCCAVFVNETAGNCPLE-----ENDVVKFYTELI-- 552  
 DB 402 NVMKDLAVHTRLTPEQORQEVGRGLDIYIHKD-DNVQRELDRDWGLSFDNSLLSFGSRIL 460  
 QY 553 ----GGCKF---RGIRIGANENRGAOSIMYDATKNEAFYKNCNTLNTGIGRFEIAATEA 604  
 DB 461 EKIHOGGKTFDYNQPADWSKETRGAPLISVKPLDNWLLIYTR-----RNYEAANSLLI 513  
 QY 605 KNMFRLP-----DK-----EOKV-----LMFIIISKRQLNAYGFVK 636  
 DB 514 QNLKPVTPAMGIQMKKAIMIEVDRTTEAYLRALQOKVTSDTQIVVCLLSNRKDKYDAIK 573  
 QY 637 HYCDHTIGVANQHITSETVTKALASLRHEKSGKEIFYOIALKINAKLGGINQELDWSEIA 696  
 DB 574 KYLCTDCTPSPQCVVARTLGK-----QOTVMAIATKIALQMNCKMGGELWRVDMA-- 623  
 QY 697 EISPEKERRKTMPLTMVYGVIDVTHPTSYSGIDYSIAAVVASINPGGTYIRNMIVTQEB 756  
 DB 624 -----LKLAMIVGIDYCHDWTAG--RRSIAGFVASINEGTRWFSRCVFQD-- 667  
 QY 757 RPERAVAHGRERTDILEAKFVKLLREFAEANDNRPAPAHIVYVRDGVSDSEMLRVSHDEL 816  
 DB 668 ----RGQELVDGLKVCLOAALRAWNSGCNE-YMPSRVIYVRDGVGDGQLKTL----- 713  
 QY 817 RSLKSEVQKQFMS--ERDGEDPEPKYTFIVIOKRHNTRLLRRMEKDKPVVVKDLTPAETD 874  
 DB 714 --VNYEVQFQFLDCLSKSGVGNPRUTVIVVKRVNTRFF----- 750  
 QY 875 AVAAVKQWEDMKESKETG--IVNPSGGTTVDKLIYSKYKDFDFLASHHGVLGTSRPGHY 932  
 DB 751 -----AQSGRLQNLPLPGTVIDVETRPEDWYDFVVSQAVRSGSVSPHY 795  
 QY 933 TVMYDDKMSQDEVKMTYGLAFLSARCKKPISLPVPVHYAH 974  
 DB 796 NVYDSSGLKPDHIORLTYKLCCHIYNNWPGVIRVPAPCOYAH 837

## RESULT 10

US-10-104-047-3439  
 ; Sequence 3439, Application US/10104047  
 ; Patent No. 6943241  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. 6943241el full length cdna  
 ; FILE REFERENCE: HI-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104,047





Db 747 -----VQNPPLGTVDSEATRNEMDYFLISQV 774

Qy 922 GVLGTSRPHGYTMVDDKMGSDQEVYKMYGLAFLSARCRKPISLPVPVHYAH 974

Db 775 ACRGTVSPYVYVDDNGLKPDHMQRLTFKLCHLYNNPFGIVSPAPQYAH 827

RESULT 12

US-09-248-796A-20922

; Sequence 20922, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 20922

; LENGTH: 257

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-20922

Query Match 5.0%; Score 265; DB 2; Length 257;

Best Local Similarity 26.0%; Pred. No. 1.7e-15;

Matches 75; Conservative 54; Mismatches 112; Indels 48; Gaps 9;

Qy 714 YGIDVTHPTSGIDYSIAAVASI-NPGGTIYRNMIVTQECRPGERAVAHGRERTDI 772

Db 6 YLVLDVT--SSQGETYSVSIAISVSGSEDGIFNKF-----PGSVRIQTGGQEV-I 54

Qy 773 LEAKFVKLLR-EPAENNDNRAPAHIVYVRDGVSDSEMLRVSHDELRLSKSEVKQFMSERD 831

Db 55 ADVKSMVLERLENFHKIKGLPSKVLFRDGVSEGGYTTILKEELTKIAAFNEYGLKN 114

Qy 832 GEDPEPKYFIVIQKRNTRLRRMEKDKPVVKNKOLTPAETDVAVAVKQWEEDMKESKE 891

Db 115 IPKYSTIIFMIVVKRHTRFI-----PIHDN-----ADDPKTKQ 150

Qy 892 TGIV---NPSSGTTVDKLVSKYKFFFLASHHGVLTGSRPGHYTVWYDDKMGSDQEVYK 948

Db 151 IAVTSNENVIAGTVDREITSPAYDFYVQSQSLQGTGIPAHYVYLHDENNYTSDTIQK 210

Qy 949 MTYGLAFLSARCRKPISLPVPVHYAHLSCAKELRYTYKEHYIGDYAQ 997

Db 211 ITYDLCHTFSRATKSVKVPVPAAYADLLCTGRD-----YIYFAK 251

RESULT 13

US-10-043-774B-13

; Sequence 13, Application US/10043774B

; Patent No. 6900017

; GENERAL INFORMATION:

; APPLICANT: University of Illinois at Chicago

; APPLICANT: Sharma, Arun

; APPLICANT: Hoffman, Ronald

; TITLE OF INVENTION: HUMAN HEMATOPOIETIC GROWTH REGULATORY GENE AND USES

; FILE REFERENCE: MBHB: CU08/PPA

; CURRENT APPLICATION NUMBER: US/10/043,774B

; CURRENT FILING DATE: 2002-06-06

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 13

; LENGTH: 844

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: PIWI protein

US-10-043-774B-13

Query Match 3.9%; Score 209.5; DB 2; Length 844;

Best Local Similarity 18.6%; Pred. No. 1.4e-09;

Matches 173; Conservative 140; Mismatches 283; Indels 335; Gaps 46;

Qy 231 DPNRFEQSLVAPRIEA-----WFGIY-----IGIKELPDGEVVL---NFAIVD 271

Db 48 EASRERRALEEAPRREGGPPKPGQDYDLNTRPVELSVKSGKDGVMQLQTNFRK 107

Qy 272 KLFYNAPKMSLDYLLLVDPQSCNDVRKDKTK---LMAGKMTIRQAARPRIRQLLE 327

Db 108 ----TKPEWRIVHYHVEP-EPSIENPRVVMGVLSNHNLLGSG-----YLPD 149

Qy 328 NLKKAECVWDNEMS-----RLTERHLTFLDLCENSL--VYKVTGKSQRGN 373

Db 150 GLQLFTTRKFEQEIIVLSGSKLDIEYKISIKFVGFISCAEPRFLQVLNLLRRSMKGLN 209

Qy 374 AKKYDTTLFKIYEENKRIEPPHLPLVKVKSACAKYAVPM-----EH-----LE 417

Db 210 LELVGRNLPD-----PRAKIE--IREFKMELWPGYETSIHQEKDILLGTE 283

Qy 418 VHEKPORVKRIDLVMQDKFLKRAKPKHDYKE---NTLKMKLKELDFSE--ELNFVER 471

Db 254 ITHKVMRTETIYDI-----MRCSHNPARHQDEVRVNVLDLVLVDYNNRTYRINDVD- 306

Qy 472 FGLCSKLQWIECPGKVL-----KEPMLVN-----495

Db 307 FQOTPK-STFSCKGRDISFVEYLYTKYINIRDHNPQLLISKNRKALKTNASELVVLIP 365

Qy 496 -----SYNEQIKMTPV-----IRGFOEKQNLNVPEKEKLCCLCAVYV 530

Db 366 ELCRVTVGLNAEMRSFQLMRAMSSYTRMNPKORTDLRAFNRLQN-TPE-----SVKVL 419

Qy 531 NETAGNPLEENDVVKFYTELIG--GCKPRGRIRIGANENRGAQSIWYDATKNEYAFYKNC 598

Db 420 RD--WNMELDKN-VTEVOGRIIGQONIYPHNGKVPAGENADWQRHFRDOR-----MLT 469

Qy 589 TLNTGIGRFEIAATE-----AKNMFERLPDKQKV-----618

Db 470 TPSDGLDRWAVIAPQRNSHELATLLDSLYRAASGMGLRIRSPQEFIIYDDRTGYVRAMD 529

Qy 619 -----LMFIIISKQLNAYFKVGH--YCDHTIGVANQHITSETVTKALASLRHEKGS 668

Db 530 DCVRSDPKLILCLVPNDNAERYSSIKRGYVDR--AVPTQVVTLKTKTKPKYSLM-----581

Qy 669 KRIFYQIALKINAKLGGINQELDWSEIAEISPEEKERRKTMPLT--MYVIGIDVTHPTSYS 736

Db 582 -SIATKIALQNLCKLG-----YTPWMI-----LPLSGLMTIGFDIAKSTRDR 623

Qy 727 GIDYSIAAVASINPGGTIYRN--MIVTOECRPGERAVAHGRERTDILE----AKFVKL 780

Db 624 KRAY--GALIASMD---LQONSTYFSTVTEC-----SAFDVLANTLWPMIAKA 666

Qy 781 LREFAENNDNRAPAHIVYVRDGVSDSEMLRVSH---DELRSIKSVKQFMSERDGEDPE 836

Db 667 LRQY-OHEHRKLPISRIVFYRDGVSGSLKQLFEFEVKDIIEKLUYARVQLS-----P 719

Qy 837 PKYTFIVIQKRNTRLRRMEKDKPVVKNKDLTPAETDVAVAVKQWEEDMKESKETGIVN 896

Db 720 PQLAIVVTRSNTRFNLNGQ-----N 741

Qy 897 PSSGTTVDKLVSKYKFFFLASHHGVLTGSRPGHYTVWYDDKMGSDQEVYKMYGLAFL 956

Db 742 PPPGTIVDDVITLPERYDFYVLSQQVROGTVSPTSYNVLYSSMGLSPEKQKQLTKYKCHL 801

Qy 957 ----SARCRKPISLPVPVHYAHLSCAKEL 983

Db 802 YYNWSGTTTRVP-----AVCOYAKKL 821



Search completed: July 5, 2006, 13:35:41  
Job time : 58 secs

---

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 13:46:11 ; Search time 189 Seconds  
(without alignments)  
2499.890 Million cell updates/sec

Title: US-10-645-746-3  
Perfect score: 5349  
Sequence: 1 MSSNFPFLEKGFYRHSILDPE.....RHEMEHPLQTNVVKYPGNSFA 1020

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5349	100.0	1020	5	US-10-645-746-3
2	5349	100.0	1020	5	US-10-645-746-3
3	5349	100.0	1020	6	US-11-144-985-3
4	4271	79.8	818	5	US-10-645-746-13
5	4271	79.8	818	5	US-10-645-746-13
6	4271	79.8	818	6	US-11-144-985-13
7	575.5	10.8	969	3	US-09-533-029-106
8	575.5	10.8	969	4	US-10-286-264-102
9	575.5	10.8	969	4	US-10-374-780A-2084
10	575.5	10.8	969	4	US-10-412-699B-752
11	566	10.6	860	6	US-10-467-397-7
12	560.5	10.5	950	6	US-11-097-143-27930
13	560.5	10.5	984	6	US-11-097-143-14511
14	560.5	10.5	984	6	US-11-097-143-27927
15	557	10.4	766	5	US-10-645-746-9
16	557	10.4	766	5	US-10-645-746-9
17	557	10.4	766	6	US-11-144-985-9
18	555.5	10.4	737	5	US-10-645-746-10
19	555.5	10.4	737	5	US-10-645-746-10
20	555.5	10.4	737	6	US-11-144-985-10
21	555.5	10.4	860	5	US-10-756-149-4706
22	555	10.4	924	4	US-10-408-765A-995
23	549.5	10.3	816	5	US-10-483-505-31
24	536.5	10.0	1034	4	US-10-437-963-192340
25	536	10.0	949	4	US-10-437-963-108712
26	533.5	10.0	979	4	US-10-437-963-121457
27	519	9.7	1048	4	US-10-174-363-56

28 519 9.7 1048 4 US-10-374-780A-2086  
29 519 9.7 1048 6 US-11-093-888-56  
30 509.5 9.5 990 4 US-10-437-963-108714  
31 507 9.5 1109 4 US-10-437-963-124320  
32 502 9.4 909 4 US-10-437-963-200613  
33 499.5 9.3 935 4 US-10-437-963-138118  
34 499.5 9.3 1088 4 US-10-437-963-158455  
35 496.5 9.3 1125 4 US-10-425-114-51644  
36 495 9.3 577 4 US-10-174-363-2  
37 495 9.3 577 6 US-11-093-888-2  
38 490 9.2 1053 4 US-10-437-963-203412  
39 489 9.1 1101 4 US-10-174-363-54  
40 489 9.1 1101 6 US-11-093-888-54  
41 488 9.1 988 4 US-10-225-066A-152  
42 488 9.1 988 4 US-10-374-780A-238  
43 488 9.1 988 5 US-10-225-066A-152  
44 487 9.1 821 4 US-10-174-363-4  
45 487 9.1 821 6 US-11-093-888-4

#### ALIGNMENTS

RESULT 1  
US-10-645-746-3  
; Sequence 3, Application US/10645746  
; Publication No. US20040265839A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabata, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; FILE REFERENCE: UMY-052DV1  
; CURRENT APPLICATION NUMBER: US/10/645,746  
; CURRENT FILING DATE: 2003-08-20  
; PRIOR APPLICATION NUMBER: US 09/689,992  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/193,218  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: US 60/159,776  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1020  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-645-746-3

Query Match 100.0%; Score 5349; DB 5; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MSSNFPFLEKGFYRHSILDPEMKWLARPTGCDGKFEKVKVLLVNVWPKFSKIYDREYVE 60  
Db 1 MSSNFPFLEKGFYRHSILDPEMKWLARPTGCDGKFEKVKVLLVNVWPKFSKIYDREYVE 60  
Qy 61 YEVMTEKVLNRKPGKPPKKTETIPDRAKLFQHLRHEKKOTDFTLEDVVPDEKDTVY 120  
Db 61 YEVMTEKVLNRKPGKPPKKTETIPDRAKLFQHLRHEKKOTDFTLEDVVPDEKDTVY 120  
Qy 121 SVCLRLNTVTSKMLVSEKVKVVKSEKDEKLEKKILYTLTYRKKFHLNFSRENPEKDE 180  
Db 121 SVCLRLNTVTSKMLVSEKVKVVKSEKDEKLEKKILYTLTYRKKFHLNFSRENPEKDE 180  
Qy 181 EANRSYKFLKNVMTQKVYAPFVNEEIKVQFAKNFYVNNNSILRVPSFHDHPRFEOSLE 240  
Db 181 EANRSYKFLKNVMTQKVYAPFVNEEIKVQFAKNFYVNNNSILRVPSFHDHPRFEOSLE 240  
Qy 241 VAPRIEAWFGIYIGIKELPDGEVPLNPAIVDKLFYNAPKMSLADYLLIIVDPQSCNDVVR 300

```
Db 241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLIIVDQSCNDDVR 300
Qy 301 KOLTKTLMAGKWTIROAARPRIRQLLENLKLKCAEVWDMESRLTERHLTFLDLCENSL 360
Db 301 KOLTKTLMAGKWTIROAARPRIRQLLENLKLKCAEVWDMESRLTERHLTFLDLCENSL 360
Qy 361 VYKVTGKSDRGNAKKYDTTLFKIYEENKKFTIEFPHLPLVKVKGAKYAVPMEHLEVHE 420
Db 361 VYKVTGKSDRGNAKKYDTTLFKIYEENKKFTIEFPHLPLVKVKGAKYAVPMEHLEVHE 420
Qy 421 KPORYKNRIDLVMQDKFLKRAIRKPHDYKENTLKMELDFSSSEELNFVERFGLCSKLM 480
Db 421 KPORYKNRIDLVMQDKFLKRAIRKPHDYKENTLKMELDFSSSEELNFVERFGLCSKLM 480
Qy 481 IECPGKVLKEPMLVNSVNEQIKMTPIRGFOEQKLNVPVPEKELCCAVFVNETAGNPCLE 540
Db 481 IECPGKVLKEPMLVNSVNEQIKMTPIRGFOEQKLNVPVPEKELCCAVFVNETAGNPCLE 540
Qy 541 ENDVVKFYTELGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600
Db 541 ENDVVKFYTELGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600
Qy 601 ATEAKMFERLPDKEQKVMFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALA 660
Db 601 ATEAKMFERLPDKEQKVMFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALA 660
Qy 661 SLRHEKSGRIFYQIALKINAKLGGINQELDSEIAEISPEEKERRKTMPLTMYGIDVT 720
Db 661 SLRHEKSGRIFYQIALKINAKLGGINQELDSEIAEISPEEKERRKTMPLTMYGIDVT 720
Qy 721 HPTSYSGIDYSIAAVASINPGGTIYRNMIVTQECRPGERAVAHGRETRDILEAKFVKL 780
Db 721 HPTSYSGIDYSIAAVASINPGGTIYRNMIVTQECRPGERAVAHGRETRDILEAKFVKL 780
Qy 781 LREFAENNDNRAPAHIVVYRDGVSDEMLRVSHDELRSLSKSVKQFMSRDEGDEPEPKYT 840
Db 781 LREFAENNDNRAPAHIVVYRDGVSDEMLRVSHDELRSLSKSVKQFMSRDEGDEPEPKYT 840
Qy 841 FVIOKHNTRLRRMEKDKPVVKNKDLTPAETDVAVAVKQWEEEDKESKETGI VNPSSG 900
Db 841 FVIOKHNTRLRRMEKDKPVVKNKDLTPAETDVAVAVKQWEEEDKESKETGI VNPSSG 900
Qy 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTVMYDDKGMQDEYVKMITYGLAFLSARC 960
Db 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTVMYDDKGMQDEYVKMITYGLAFLSARC 960
Qy 961 RKPISLPVPVHYAHLSCAKELRYTYKEHYIGDYAQPRTREHMEHFLQTNVKYQGMSPA 1020
Db 961 RKPISLPVPVHYAHLSCAKELRYTYKEHYIGDYAQPRTREHMEHFLQTNVKYQGMSPA 1020

RESULT 2
US-10-645-735-3
; Sequence 3, Application US/10645735
; Publication No. US20050100913A1
; GENERAL INFORMATION:
; APPLICANT: Mello, Craig C.
; APPLICANT: Tabara, Hiroaki
; APPLICANT: Grishok, Alla
; APPLICANT: Fire, Andrew
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC
; TITLE OF INVENTION: INTERFERENCE
; FILE REFERENCE: UMG-052
; CURRENT APPLICATION NUMBER: US/10/645,735
; CURRENT FILING DATE: 2003-08-20
; PRIOR APPLICATION NUMBER: US/09/689,992A
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 3
; LENGTH: 1020
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-645-735-3

Query Match      100.0%; Score 5349; DB 5; Length 1020;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSSNPPELEKGFYRHSLDPEMKWMLARPTGKCDGKFYEKKVLLVMNFKSSKIYDREYVE 60
Db 1 MSSNPPELEKGFYRHSLDPEMKWMLARPTGKCDGKFYEKKVLLVMNFKSSKIYDREYVE 60
Qy 61 YEYKMTKEVLNRKPKGPPPKTEIIPDRAKLFWOHLRHEKKQTDPILEDYVDEKDTVY 120
Db 61 YEYKMTKEVLNRKPKGPPPKTEIIPDRAKLFWOHLRHEKKQTDPILEDYVDEKDTVY 120
Qy 121 SVCRLNTVTSKMLVSEKVVVKDSEKKDEKLEKKILYTMILTYRKKFHLNFSRENPEKDE 180
Db 121 SVCRLNTVTSKMLVSEKVVVKDSEKKDEKLEKKILYTMILTYRKKFHLNFSRENPEKDE 180
Qy 181 EANRSYKFLKNVMTQKVYAPFVNBEIKVQFAKNFYDNNSILRVPESEPHDPNRFQOSLE 240
Db 181 EANRSYKFLKNVMTQKVYAPFVNBEIKVQFAKNFYDNNSILRVPESEPHDPNRFQOSLE 240
Qy 241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLIIVDQSCNDDVR 300
Db 241 VAPRIEAWFGIYIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLLLIIVDQSCNDDVR 300
Qy 301 KOLTKTLMAGKWTIROAARPRIRQLLENLKLKCAEVWDMESRLTERHLTFLDLCENSL 360
Db 301 KOLTKTLMAGKWTIROAARPRIRQLLENLKLKCAEVWDMESRLTERHLTFLDLCENSL 360
Qy 361 VYKVTGKSDRGNAKKYDTTLFKIYEENKKFTIEFPHLPLVKVKGAKYAVPMEHLEVHE 420
Db 361 VYKVTGKSDRGNAKKYDTTLFKIYEENKKFTIEFPHLPLVKVKGAKYAVPMEHLEVHE 420
Qy 421 KPORYKNRIDLVMQDKFLKRAIRKPHDYKENTLKMELDFSSSEELNFVERFGLCSKLM 480
Db 421 KPORYKNRIDLVMQDKFLKRAIRKPHDYKENTLKMELDFSSSEELNFVERFGLCSKLM 480
Qy 481 IECPGKVLKEPMLVNSVNEQIKMTPIRGFOEQKLNVPVPEKELCCAVFVNETAGNPCLE 540
Db 481 IECPGKVLKEPMLVNSVNEQIKMTPIRGFOEQKLNVPVPEKELCCAVFVNETAGNPCLE 540
Qy 541 ENDVVKFYTELGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600
Db 541 ENDVVKFYTELGGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCITLNTGIGRFEIA 600
Qy 601 ATEAKMFERLPDKEQKVMFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALA 660
Db 601 ATEAKMFERLPDKEQKVMFIIISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALA 660
Qy 661 SLRHEKSGRIFYQIALKINAKLGGINQELDSEIAEISPEEKERRKTMPLTMYGIDVT 720
Db 661 SLRHEKSGRIFYQIALKINAKLGGINQELDSEIAEISPEEKERRKTMPLTMYGIDVT 720
Qy 721 HPTSYSGIDYSIAAVASINPGGTIYRNMIVTQECRPGERAVAHGRETRDILEAKFVKL 780
Db 721 HPTSYSGIDYSIAAVASINPGGTIYRNMIVTQECRPGERAVAHGRETRDILEAKFVKL 780
Qy 781 LREFAENNDNRAPAHIVVYRDGVSDEMLRVSHDELRSLSKSVKQFMSRDEGDEPEPKYT 840
Db 781 LREFAENNDNRAPAHIVVYRDGVSDEMLRVSHDELRSLSKSVKQFMSRDEGDEPEPKYT 840
Qy 841 FVIOKHNTRLRRMEKDKPVVKNKDLTPAETDVAVAVKQWEEEDKESKETGI VNPSSG 900
Db 841 FVIOKHNTRLRRMEKDKPVVKNKDLTPAETDVAVAVKQWEEEDKESKETGI VNPSSG 900
Qy 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTVMYDDKGMQDEYVKMITYGLAFLSARC 960
Db 901 TTVDKLIIVSKYKFDFFLASHHGVLGTSRPGHYTVMYDDKGMQDEYVKMITYGLAFLSARC 960
```

Qy 961 RKPISLPVPHVYAHLSCEKAKELYRTYKEHYIGDYAQPRTHEHEHFLQTNVKYPGMSFA 1020  
Db 961 RKPISLPVPHVYAHLSCEKAKELYRTYKEHYIGDYAQPRTHEHEHFLQTNVKYPGMSFA 1020

RESULT 3  
US-11-144-985-3  
; Sequence 3, Application US/11144985  
; Publication No. US20060024798A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabata, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; FILE REFERENCE: UMY-052CN  
; CURRENT APPLICATION NUMBER: US/11/144,985  
; PRIOR FILING DATE: 2005-06-03 09/689,992  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1020  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-11-144-985-3

Query Match 100.0%; Score 5349; DB 6; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSSNPFLEKGFYRHSIDPEMKLARPTGKCDGKFEYKQVLLVNNPKSSKIYDREYE 60  
Db 1 MSSNPFLEKGFYRHSIDPEMKLARPTGKCDGKFEYKQVLLVNNPKSSKIYDREYE 60

Qy 61 YEYKMTKEVLRKPGFPKPTKIPIDRAKLFQHLRHEKKTDFILEDYVDEKDTVY 120  
Db 61 YEYKMTKEVLRKPGFPKPTKIPIDRAKLFQHLRHEKKTDFILEDYVDEKDTVY 120

Qy 121 SVCLRLNTVTSKMLVSEKVVKDEKDEKLEKKILYTLTYRKKFPHLNFSPENPEKDE 180  
Db 121 SVCLRLNTVTSKMLVSEKVVKDEKDEKLEKKILYTLTYRKKFPHLNFSPENPEKDE 180

Qy 181 EANSYKFLQNVMTQKRYAPFVNEBIKVQFKNFVYDNNISILRVPSFHDPPNRFQSL 240  
Db 181 EANSYKFLQNVMTQKRYAPFVNEBIKVQFKNFVYDNNISILRVPSFHDPPNRFQSL 240

Qy 241 VAPRIEAWGIGYIGIELEGEVPLNFAIVDKLPYNAKMSLLDYLLIIVDPQSCNDV 300  
Db 241 VAPRIEAWGIGYIGIELEGEVPLNFAIVDKLPYNAKMSLLDYLLIIVDPQSCNDV 300

Qy 301 KDLTKLMAGMTIRQAARPIRQLLENLKLCAEVDNEMSLTERHLTFLDLCENSL 360  
Db 301 KDLTKLMAGMTIRQAARPIRQLLENLKLCAEVDNEMSLTERHLTFLDLCENSL 360

Qy 361 VYKVGKSGRGNKAYDTTLTKIYBENKFIFFPHLPLVKVSGAKAYAVPMHELVHE 420  
Db 361 VYKVGKSGRGNKAYDTTLTKIYBENKFIFFPHLPLVKVSGAKAYAVPMHELVHE 420

Qy 421 KPORYKNRIDLVMQDFELKRAATKPHDYKENTLKMKELDFFSSEELNFVERFGLCSK 480  
Db 421 KPORYKNRIDLVMQDFELKRAATKPHDYKENTLKMKELDFFSSEELNFVERFGLCSK 480

Qy 481 IECPCGKVLKPEMLVNSVNEQIKMTPIRVGFQEKQNVPEKELCCAVFVNNETAGN 540  
Db 481 IECPCGKVLKPEMLVNSVNEQIKMTPIRVGFQEKQNVPEKELCCAVFVNNETAGN 540

Qy 541 ENDVVKFYTELIGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCCTLNTGIRPEIA 600  
Db 541 ENDVVKFYTELIGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCCTLNTGIRPEIA 600

Qy 601 ATEAKMFERLPDKEQKVLMFIIISKQLANAYGVKHVCDHTIGVANOHITSETVTKALA 660  
Db 601 ATEAKMFERLPDKEQKVLMFIIISKQLANAYGVKHVCDHTIGVANOHITSETVTKALA 660

Qy 661 SLRHEKSGKRFYQIALKINAKLGGINOELDWSEIAEISPEEKERRKTMPLTMVVGIDVT 720  
Db 661 SLRHEKSGKRFYQIALKINAKLGGINOELDWSEIAEISPEEKERRKTMPLTMVVGIDVT 720

Qy 721 HPTSYSGIDYSIAAVASINPGGTIVRNMTVTOEECPGGERAVAHGERTDILBAKPVKL 780  
Db 721 HPTSYSGIDYSIAAVASINPGGTIVRNMTVTOEECPGGERAVAHGERTDILBAKPVKL 780

Qy 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLVSHDELRLSLSEVKQFMSERDGEDPEPKYT 840  
Db 781 LREFAENNDNRAPAHIVVYRDGVSDSEMLVSHDELRLSLSEVKQFMSERDGEDPEPKYT 840

Qy 841 FIVIQKHNTLRRLRMEKDKPVVVKDLPATDVAVAQVQWEDMKESKETGIVNPSG 900  
Db 841 FIVIQKHNTLRRLRMEKDKPVVVKDLPATDVAVAQVQWEDMKESKETGIVNPSG 900

Qy 901 TTVDKLIIVSKYKDFFLASHHGVLTSRPGHYTVMYDDKQMSQDEVYKMTYGLAFLSARC 960  
Db 901 TTVDKLIIVSKYKDFFLASHHGVLTSRPGHYTVMYDDKQMSQDEVYKMTYGLAFLSARC 960

Qy 961 RKPISLPVPHVYAHLSCEKAKELYRTYKEHYIGDYAQPRTHEHEHFLQTNVKYPGMSFA 1020  
Db 961 RKPISLPVPHVYAHLSCEKAKELYRTYKEHYIGDYAQPRTHEHEHFLQTNVKYPGMSFA 1020

## RESULT 4

US-10-645-746-13  
; Sequence 13, Application US/10645746  
; Publication No. US20040265839A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabata, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; FILE REFERENCE: UMY-052DV1  
; CURRENT APPLICATION NUMBER: US/10/645,746  
; PRIOR FILING DATE: 2003-08-20  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 818  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-645-746-13

Query Match 79.8%; Score 4271; DB 5; Length 818;  
Best Local Similarity 100.0%; Pred. No. 4,7e-316;  
Matches 818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 VNEEIKVQPAKPVYDNNISILRVPSFHDPPNRFQSLVAPRIEAFVGIYIGIKELPDGE 262  
Db 1 VNEEIKVQPAKPVYDNNISILRVPSFHDPPNRFQSLVAPRIEAFVGIYIGIKELPDGE 60

Qy 263 PVLNFAIVDKLPYNAKMSLLDYLLIIVDPQSCNDVDRKDLTKLMAGKWTIRQAARPI 322  
Db 61 PVLNFAIVDKLPYNAKMSLLDYLLIIVDPQSCNDVDRKDLTKLMAGKWTIRQAARPI 120

```
QY 323 ROLLENLKLKCAEVDNEMSLRTERHLTFDLCEENSLVYKVTGKSDRGNAKKYDTTLF 382
Db 121 ROLLENLKLKCAEVDNEMSLRTERHLTFDLCEENSLVYKVTGKSDRGNAKKYDTTLF 180
QY 383 KIYEENKKTIEPPLPLVKVKSAGKEYAVPMEHLEVHEKPPQRYKNRIDLVMDQKFLKRAT 442
Db 181 KIYEENKKTIEPPLPLVKVKSAGKEYAVPMEHLEVHEKPPQRYKNRIDLVMDQKFLKRAT 240
QY 443 RPHDYKENTLKMKELDSSSEELNVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 502
Db 241 RPHDYKENTLKMKELDSSSEELNVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 300
QY 503 MTPVIRGFOEKQNLNVPEKELCCAVFVNENAGNCPCLLENDVVKFTELIIGCKRGIRI 562
Db 301 MTPVIRGFOEKQNLNVPEKELCCAVFVNENAGNCPCLLENDVVKFTELIIGCKRGIRI 360
QY 563 GANENRGAQSIIMYDATKNEYAFYKNCITLNTGIRFEIAATEAKNMFERLPDKEQKVLMEI 622
Db 361 GANENRGAQSIIMYDATKNEYAFYKNCITLNTGIRFEIAATEAKNMFERLPDKEQKVLMEI 420
QY 623 IISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALASLRHEKSGKRIFYQIALKINAK 682
Db 421 IISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALASLRHEKSGKRIFYQIALKINAK 480
QY 683 LGGINQELDWEISAEISPEEKERRKTMTPLTMVVGIDVTHPTSYSGIDYSIAAVASINPG 742
Db 481 LGGINQELDWEISAEISPEEKERRKTMTPLTMVVGIDVTHPTSYSGIDYSIAAVASINPG 540
QY 743 GTIYRNMIVTQECRPGERAVAHGRERTDILEAKFVKLLRFEAENNDNRAPAHIVVYRDG 802
Db 541 GTIYRNMIVTQECRPGERAVAHGRERTDILEAKFVKLLRFEAENNDNRAPAHIVVYRDG 600
QY 803 VSDSEMLRVSHDELRLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 862
Db 601 VSDSEMLRVSHDELRLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 660
QY 863 VNKDLTPAETDVAAVAVKQWEEDMKESKETGIVNPSGGTTVDKLIIVSKYKFPFFLASHHG 922
Db 661 VNKDLTPAETDVAAVAVKQWEEDMKESKETGIVNPSGGTTVDKLIIVSKYKFPFFLASHHG 720
QY 923 VLGTSRPGHYTMVDDKGMQSDQEVYKMTYGLAFLSARCRKPIISLPVPHYAHLSCEKAKE 982
Db 721 VLGTSRPGHYTMVDDKGMQSDQEVYKMTYGLAFLSARCRKPIISLPVPHYAHLSCEKAKE 780
QY 983 LYRTYKHYIGDYAOPRTRHEMEHFLQTNVVKPGMSFA 1020
Db 781 LYRTYKHYIGDYAOPRTRHEMEHFLQTNVVKPGMSFA 818

RESULT 5
US-10-645-735-13
; Sequence 13, Application US/10645735
; Publication NO. US20050100913A1
; GENERAL INFORMATION:
; APPLICANT: Mello, Craig C.
; APPLICANT: Tabara, Hiroaki
; APPLICANT: Grishok, Alla
; APPLICANT: Fire, Andrew
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC
; TITLE OF INVENTION: INTERFERENCE
; FILE REFERENCE: UMG-052
; CURRENT APPLICATION NUMBER: US/10/645,735
; CURRENT FILING DATE: 2003-08-20
; PRIOR APPLICATION NUMBER: US/09/689,992A
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/193,218
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/159,776
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
```

```
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-645-735-13

Query Match 79.8%; Score 4271; DB 5; Length 818;
Best Local Similarity 100.0%; Pred. No. 4.7e-316;
Matches 818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 VNEEIKVQPAKNFVYDNNNSILRVPSFHDNPRFEQSLVAPRIEAWFGIYIGIKELFDGE 262
Db 1 VNEEIKVQPAKNFVYDNNNSILRVPSFHDNPRFEQSLVAPRIEAWFGIYIGIKELFDGE 60
QY 263 PVLNFAIVDKLPYNAPKMSLLDYLLIIVDPQSCNDVDRKDLTKLMAGKWTIRQAARPI 322
Db 61 PVLNFAIVDKLPYNAPKMSLLDYLLIIVDPQSCNDVDRKDLTKLMAGKWTIRQAARPI 120
QY 323 ROLLENLKLKCAEVDNEMSLRTERHLTFDLCEENSLVYKVTGKSDRGNAKKYDTTLF 382
Db 121 ROLLENLKLKCAEVDNEMSLRTERHLTFDLCEENSLVYKVTGKSDRGNAKKYDTTLF 180
QY 383 KIYEENKKTIEPPLPLVKVKSAGKEYAVPMEHLEVHEKPPQRYKNRIDLVMDQKFLKRAT 442
Db 181 KIYEENKKTIEPPLPLVKVKSAGKEYAVPMEHLEVHEKPPQRYKNRIDLVMDQKFLKRAT 240
QY 443 RPHDYKENTLKMKELDSSSEELNVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 502
Db 241 RPHDYKENTLKMKELDSSSEELNVERFGLCSKLQMIETCPGKVLKEPMLVNSVNEQIK 300
QY 503 MTPVIRGFOEKQNLNVPEKELCCAVFVNENAGNCPCLLENDVVKFTELIIGCKRGIRI 562
Db 301 MTPVIRGFOEKQNLNVPEKELCCAVFVNENAGNCPCLLENDVVKFTELIIGCKRGIRI 360
QY 563 GANENRGAQSIIMYDATKNEYAFYKNCITLNTGIRFEIAATEAKNMFERLPDKEQKVLMEI 622
Db 361 GANENRGAQSIIMYDATKNEYAFYKNCITLNTGIRFEIAATEAKNMFERLPDKEQKVLMEI 420
QY 623 IISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALASLRHEKSGKRIFYQIALKINAK 682
Db 421 IISKROLNAYGFVKHYCDHTIGVANQHITSETVTKALASLRHEKSGKRIFYQIALKINAK 480
QY 683 LGGINQELDWEISAEISPEEKERRKTMTPLTMVVGIDVTHPTSYSGIDYSIAAVASINPG 742
Db 481 LGGINQELDWEISAEISPEEKERRKTMTPLTMVVGIDVTHPTSYSGIDYSIAAVASINPG 540
QY 743 GTIYRNMIVTQECRPGERAVAHGRERTDILEAKFVKLLRFEAENNDNRAPAHIVVYRDG 802
Db 541 GTIYRNMIVTQECRPGERAVAHGRERTDILEAKFVKLLRFEAENNDNRAPAHIVVYRDG 600
QY 803 VSDSEMLRVSHDELRLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 862
Db 601 VSDSEMLRVSHDELRLSKSEVKQFMSERDGEDPEPKYTFIVIQKRHNTLLRRMEKDKPV 660
QY 863 VNKDLTPAETDVAAVAVKQWEEDMKESKETGIVNPSGGTTVDKLIIVSKYKFPFFLASHHG 922
Db 661 VNKDLTPAETDVAAVAVKQWEEDMKESKETGIVNPSGGTTVDKLIIVSKYKFPFFLASHHG 720
QY 923 VLGTSRPGHYTMVDDKGMQSDQEVYKMTYGLAFLSARCRKPIISLPVPHYAHLSCEKAKE 982
Db 721 VLGTSRPGHYTMVDDKGMQSDQEVYKMTYGLAFLSARCRKPIISLPVPHYAHLSCEKAKE 780
QY 983 LYRTYKHYIGDYAOPRTRHEMEHFLQTNVVKPGMSFA 1020
Db 781 LYRTYKHYIGDYAOPRTRHEMEHFLQTNVVKPGMSFA 818

RESULT 6
US-11-144-985-13
; Sequence 13, Application US/11144985
; Publication NO. US20060024798A1
; GENERAL INFORMATION:
; APPLICANT: Mello, Craig C.
; APPLICANT: Tabara, Hiroaki
```



APPLICANT: Grishok, Alla  
APPLICANT: Fire, Andrew  
TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
TITLE OF INVENTION: INTERFERENCE  
FILE REFERENCE: UMY-052CN  
CURRENT APPLICATION NUMBER: US/11/144,985  
CURRENT FILING DATE: 2005-06-03  
PRIOR APPLICATION NUMBER: US 09/589,992  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: US 60/193,218  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: US 60/159,776  
PRIOR FILING DATE: 1999-10-15  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 818  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-11-144-985-13

Query Match 79.8%; Score 4271; DB 6; Length 818;  
Best Local Similarity 100.0%; Pred. No. 4.7e-316;  
Matches 818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 VNEEIKVQAKNFYVDNNSILRVDESFDHDPNRFQSLVAPRIEAWFGIYIGIKELFDGE 262  
Db 1 VNEEIKVQAKNFYVDNNSILRVDESFDHDPNRFQSLVAPRIEAWFGIYIGIKELFDGE 60

Qy 263 PVLNFAIVDKLFYNAPKMSLLDYLIVDPQSCNDVDRKDKTKLMAGKMTIRQAAPRI 322  
Db 61 PVLNFAIVDKLFYNAPKMSLLDYLIVDPQSCNDVDRKDKTKLMAGKMTIRQAAPRI 120

Qy 323 ROLLENLKLCAEVDNEMSLRTERHLLTFLDLCEENSLVYKVTGSDRGNAKYYDTLTF 382  
Db 121 ROLLENLKLCAEVDNEMSLRTERHLLTFLDLCEENSLVYKVTGSDRGNAKYYDTLTF 180

Qy 383 KIYEENKKTFFPHPLPVKVGSKGKEYAVPMHEHLEVEHEKQRYKNRIDLVMQDKFLKRA 442  
Db 181 KIYEENKKTFFPHPLPVKVGSKGKEYAVPMHEHLEVEHEKQRYKNRIDLVMQDKFLKRA 240

Qy 443 RKPHDYKENTLKMELKELDFSEELNLFVERFGLCSKLMICPGKVLKEPMLVNSVNOIK 502  
Db 241 RKPHDYKENTLKMELKELDFSEELNLFVERFGLCSKLMICPGKVLKEPMLVNSVNOIK 300

Qy 503 MTPVIRGFBQKQLNVVPEKELCCAVFVYVNETAGNCPLEENDVVKFYTELIGGCKFRGIRI 562  
Db 301 MTPVIRGFBQKQLNVVPEKELCCAVFVYVNETAGNCPLEENDVVKFYTELIGGCKFRGIRI 360

Qy 563 GANENRGAQSIMYDATKNEYAFYKNCYTLNTGIGRFEIATAEAKNMFRLPDKEQKVLMPFI 622  
Db 361 GANENRGAQSIMYDATKNEYAFYKNCYTLNTGIGRFEIATAEAKNMFRLPDKEQKVLMPFI 420

Qy 623 IISKROLNAGFYVKHYCDHTIGVANQHTTSETVTTKALASLRHEKSGKRIFYQIALKINAK 682  
Db 421 IISKROLNAGFYVKHYCDHTIGVANQHTTSETVTTKALASLRHEKSGKRIFYQIALKINAK 480

Qy 683 LGGINQELDSWSEIAEISPEEKERRKTMPLTWYGVIDVTHPTSYSGIDYSIAAVVASINPG 742  
Db 481 LGGINQELDSWSEIAEISPEEKERRKTMPLTWYGVIDVTHPTSYSGIDYSIAAVVASINPG 540

Qy 743 GTTYRNMIVTOEBCRPOGERAVAGRERTDILEAKFVKLLREFAENNDRAPAHIVVYRDG 802  
Db 541 GTTYRNMIVTOEBCRPOGERAVAGRERTDILEAKFVKLLREFAENNDRAPAHIVVYRDG 600

Qy 803 VDSSEMLRVSHDELRSLSKSVKQFMSERDGEDPEPKYTFIVIOKRNTLRLREWEKDKPV 862  
Db 601 VDSSEMLRVSHDELRSLSKSVKQFMSERDGEDPEPKYTFIVIOKRNTLRLREWEKDKPV 660

Qy 863 VNKDLTPAETDVAVAQWKEEDMKESKETGIVNPSGGTTVDKLIYSKYKFDFFLASHHG 922  
Db 661 VNKDLTPAETDVAVAQWKEEDMKESKETGIVNPSGGTTVDKLIYSKYKFDFFLASHHG 720

Qy 923 VLGTSRPGHYTVMYDDKMGMSODEVYKMTYGLAFISARCRKDISLPVPHVYAHLSCEKAKE 982  
Db 721 VLGTSRPGHYTVMYDDKMGMSODEVYKMTYGLAFISARCRKDISLPVPHVYAHLSCEKAKE 780

Qy 983 LYRTYKEHYIGDYAQPRTREHMEHFLQTNVVKYPGMSFA 1020  
Db 781 LYRTYKEHYIGDYAQPRTREHMEHFLQTNVVKYPGMSFA 818

RESULT 7  
US-09-533-029-106  
Sequence 106 Application US/09533029  
Publication No. US20030046723A1  
GENERAL INFORMATION:  
APPLICANT: Heard, Jacqueline  
APPLICANT: Broun, Pierre  
APPLICANT: Riechmann, Jose-Luis  
APPLICANT: Keddle, James  
APPLICANT: Pineda, Omaira  
APPLICANT: Adam, Luc  
APPLICANT: Samaha, Raymond  
APPLICANT: Zhang, James  
APPLICANT: Yu, Guo-Liang  
APPLICANT: Ratcliffe, Oliver  
APPLICANT: Pilgrim, Marsha  
APPLICANT: Jiang, Cai-Zhong  
APPLICANT: Reuber, Lynne  
TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES  
FILE REFERENCE: MBI-010  
CURRENT APPLICATION NUMBER: US/09/533,029  
CURRENT FILING DATE: 2000-03-22  
EARLIER APPLICATION NUMBER: 60/125,814  
EARLIER FILING DATE: 1999-03-23  
NUMBER OF SEQ ID NOS: 121  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 106  
LENGTH: 969  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: G1149  
US-09-533-029-106

Query Match 10.8%; Score 575.5; DB 3; Length 969;  
Best Local Similarity 23.2%; Pred. No. 2.5e-34;  
Matches 235; Conservative 161; Mismatches 339; Indels 279; Gaps 40;

Qy 38 KKVLLLVNWKFSKSIYDREYVEYKMTYKVLNRKPKGKPPKTEIPIDRAKLFQHL 97  
Db 158 KKVVRAN--HFLVQVADRDLHYDVSINPEVSK----- 190

Qy 98 RHEKKQTDFTLEDYVDFEKOVTYVSVCLNTVTGKMLVSEKVKVKDS--KDEKDLKK 154  
Db 191 -----TVNRNVKLLVN---YKDSHLGKSPAYD-GRK 220

Qy 155 ILVTM--ILTYRKXKFLNFSRENPEKDEANRSYK- LKNV-----MTQKVRYA 200  
Db 221 SLYTAGPLPDSKEFVYVNLAEKADGSGGKDRPKVAVKNTVTDLYLQQLDRLKQREA 280

Qy 201 PFVNEIKVQAKNFYVDNNSILRVDESFDHDPNRFQSL-----EVAPRIEAWFGI 251  
Db 281 PY--DTIQVLDVVLRDKPSNDYVSVGRSF-----PHTSLGKDARDGREGELGIEYWRGY 333

Qy 252 YIGIKELFDGEPVLNFAIVDKLFYNAPKMSLLDYLIVDPQSCNDVDRKDKTKLMAGK 311  
Db 334 FQSLR-LTQWGLSLNLDVSARSFY-----EPYVTDIFISFLNIRDL--N 375

Qy 312 MTRQARPRIROLLBNLKLKCAEVDNEMSLRTERHLLTFLDLCEENSLVYKVTGSD-- 369  
Db 376 RPLRDSRLKVKVKVRLTKVLLH--NNGTKSA-----KISGISLP 415

Qy 370 ----RGNNAKYYDTLTFKYEENKKTFFPHPLPVKVGSKGKEYAVPMHEHLEVEHEKQ 424  
Db 370 ----RGNNAKYYDTLTFKYEENKKTFFPHPLPVKVGSKGKEYAVPMHEHLEVEHEKQ 424

```
Db 416 IRELREFTLEDKSEKTVVQYFAEKYNYRVKYQALPAIQGSDTRPVVLPMLQCIDE--GQR 474
Qy 425 YKNRIDLVMQDFLKRATRKPHDYKENTLMLKELDFSSSEELNFRFGLCSKLQWIECP 484
Db 475 YTKRLNEKQVALLKATCORPPD--RENSIKNLVVKNNYNDLLS--KEFGMSVTTQLASIE 531
Qy 485 GKVLKEPMLVNSNEQIKWTPVIRGQEKQLNVPEKELCCAVFVNVNAGNCPLENDV 544
Db 532 ARVLPPMLKYHDSGKEMVNPRLG---QWNNIDKK----- 564
Qy 545 VRPEYTELIGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCNTLNTGIGRFEIAATEA 604
Db 565 -----MVNGAK---VTSWTCFEPQPAIPFISCPPEH-----IEEALLD- 600
Qy 605 KMFERLPDKEQKVLNFIISKRQLNAYGFVKYCHDHTIGVANQHITSETVTKALASLRH 664
Db 601 --IHKEAPGLQ---LLVILPDVTGSGYKIKRICETELGIVSQCCQPRQVKNL----- 648
Qy 665 EKGSKRIFYQIAIKNAKLGGINOELDWSEIAEISPEEKERRKTMPL-----TMYVGIDV 719
Db 649 ---NKQYMENVALKINVTGGRTVLN-----DAIRRNIPLIITDRPTIINGADV 694
Qy 720 THPTSYSGIDYSIAAVVASIN--PGGTIYRNMIWTO---EE-----CRPGERAVAHGR 767
Db 695 THPQGEDSSPSIAAVVASMDWPEINKYGLVSAQAHRHEIIQDLKLVQDPQGLVH-- 752
Qy 768 ERTDILEAKFVKLLREFAEENNDRAPAHIVVYRDGVSDSEMLRVSHDELSLKSEVKQFM 827
Db 694 ---SGLIREHFIAPFRA-----TGQIPQRIIFYRDGVSEGFQSVLLHEMTAIRKACNSLQ 805
Qy 828 SERDGEDPPEPKYTFVIOQRHNTRLRLRMKDKPVVKNKDLTPAETDVAVAQWEEDEMK 887
Db 806 -----ENYVPRVTFVIOQRHTRLFPBOHG-----NRDMT----- 836
Qy 888 ESKETGINVPSGTTVDKLVISKYKDFPLASHHGVLGTSRPGHYTYVMYDDKMSQDEVY 947
Db 837 ---DKSGNIQP--GTVVDTKICHNEFDFFYLNHAGIQGTSRPAHYHVLDDENGFTADQLQ 892
Qy 948 KMTYGLAFLSARCRKPISLPVPVHYAHLSCAKELYRKYKEHYIGDYAQPRTR 1001
Db 893 MLTNNLCYTYARCTKSIVPPPAYAHAAFPRA---RYYMESEMSDGGSSRSR 942
```

## RESULT 8

US-10-286-264-102

; Sequence 102, Application US/10286264

; Publication No. US20030093837A1

## GENERAL INFORMATION:

; APPLICANT: Keddie, James

; APPLICANT: Riechmann, Jose-Luis

; APPLICANT: Ratcliffe, Oliver

; APPLICANT: Zhang, James

; APPLICANT: Jiang, Cai-Zhong

; APPLICANT: Pineda, Omaira

; APPLICANT: Heard, Jacqueline

; APPLICANT: Yu, Guo-Liang

; APPLICANT: Adam, Luc

; APPLICANT: Broun, Pierre

; APPLICANT: Reuber, Lynne

; APPLICANT: Pilgrim, Marsha

; APPLICANT: Samaha, Raymond

; TITLE OF INVENTION: POLYNUCLEOTIDES FOR SEED TRAIT ALTERATION

; FILE REFERENCE: MBI-008

; CURRENT APPLICATION NUMBER: US/10/286,264

; CURRENT FILING DATE: 2002-11-01

; PRIOR APPLICATION NUMBER: 60/125,814

; PRIOR FILING DATE: 1999-03-23

; NUMBER OF SEQ ID NOS: 165

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 102

; LENGTH: 969

; TYPE: PRN

; ORGANISM: Arabidopsis thaliana

; FEATURE:  
; OTHER INFORMATION: G1149

US-10-286-264-102

Query Match 10.8%; Score 575.5; DB 4; Length 969;  
Best Local Similarity 23.2%; Pred. No. 2.5e-34;  
Matches 235; Conservative 161; Mismatch 339; Indels 279; Gaps 40;

```
Qy 38 KVVLLVNWFKSSKIYDREYVEYVQMTKEVNLKPKGPKFPKTEIPIPDRAKLFWHL 97
Db 158 KKVVRAN--HFLVQVADRDLHYDVSNPEISK----- 190
Qy 98 RHEKQTDFILEDDYVDEKDTVYVCRNLNTVTSKMLVSEKVVKKDSE---KKDEKLEKK 154
Db 191 -----TVNRNVKLLVN---YKDSHLGGKSPAYD--GRK 220
Qy 155 ILYTM--ILTVRKPKFLNFSRENPEKDEBANRSYKF--LQNV-----MTQKRVYA 200
Db 221 SLYTAGPLPFDSEKFVWNLAEKRADGSSGKDRPFKVAVKNTVSTDLVYQLQQFLDKRQEA 280
Qy 201 PFVNEEIKVQFAKNFYVDNNSILRVPESEPHDPRFEQSL-----EVAPRIEAWFGI 251
Db 281 PY--DTIQVLDVVLDRKPSNDYVSGRSP-----PHTSLGKDARDGRGLGDIYWRGY 333
Qy 252 YIGIKELPDGEPVLFNFAIVDKLFYNAPKMSLLDYLILLIVDPOSNDNDVAKDKTKLMAGK 311
Db 334 FOSIL--LTQMGLSLNIDVSARSFY-----EPIVVTDFISKFLNIRD--N 375
Qy 312 MTIROAARPRIQLLENLKLKCAEVDNEMSLRTERHLTFDLCEENSIVYKVTGKSD-- 369
Db 376 RPLRSDRLKVKVKVLRLLTKVLLH--WNGTKSA-----KISGISLUP 415
Qy 370 ----RGRNAKKYDTTLFKIYEENKCF--IEFPHLPLVKVKSGAKAYAVPMHELVHEKPKR 424
Db 416 IRELFTLEDKSEKTVVQYFAEKYNYRVKYQALPAIQGSDTRPVVLPMLQCIDE--GQR 474
Qy 425 YKNRIDLVMQDFLKRATRKPHDYKENTLMLKELDFSSSEELNFRFGLCSKLQWIECP 484
Db 475 YTKRLNEKQVALLKATCORPPD--RENSIKNLVVKNNYNDLLS--KEFGMSVTTQLASIE 531
Qy 485 GKVLKEPMLVNSNEQIKWTPVIRGQEKQLNVPEKELCCAVFVNVNAGNCPLENDV 544
Db 532 ARVLPPMLKYHDSGKEMVNPRLG---QWNNIDKK----- 564
Qy 545 VKFYTELIGCKFRGIRIGANENRGAQSIMYDATKNEYAFYKNCNTLNTGIGRFEIAATEA 604
Db 565 -----MVNGAK---VTSWTCFEPQPAIPFISCPPEH-----IEEALLD- 600
Qy 605 KMFERLPDKEQKVLNFIISKRQLNAYGFVKYCHDHTIGVANQHITSETVTKALASLRH 664
Db 601 --IHKEAPGLQ---LLVILPDVTGSGYKIKRICETELGIVSQCCQPRQVKNL----- 648
Qy 665 EKGSKRIFYQIAIKNAKLGGINOELDWSEIAEISPEEKERRKTMPL-----TMYVGIDV 719
Db 649 ---NKQYMENVALKINVTGGRTVLN-----DAIRRNIPLIITDRPTIINGADV 694
Qy 720 THPTSYSGIDYSIAAVVASIN--PGGTIYRNMIWTO---EE-----CRPGERAVAHGR 767
Db 695 THPQGEDSSPSIAAVVASMDWPEINKYGLVSAQAHRHEIIQDLKLVQDPQGLVH-- 752
Qy 768 ERTDILEAKFVKLLREFAEENNDRAPAHIVVYRDGVSDSEMLRVSHDELSLKSEVKQFM 827
Db 753 ---SGLIREHFIAPFRA-----TGQIPQRIIFYRDGVSEGFQSVLLHEMTAIRKACNSLQ 805
Qy 828 SERDGEDPPEPKYTFVIOQRHNTRLRLRMKDKPVVKNKDLTPAETDVAVAQWEEDEMK 887
Db 806 -----ENYVPRVTFVIOQRHTRLFPBOHG-----NRDMT----- 836
Qy 888 ESKETGINVPSGTTVDKLVISKYKDFPLASHHGVLGTSRPGHYTYVMYDDKMSQDEVY 947
Db 837 ---DKSGNIQP--GTVVDTKICHNEFDFFYLNHAGIQGTSRPAHYHVLDDENGFTADQLQ 892
Qy 948 KMTYGLAFLSARCRKPISLPVPVHYAHLSCAKELYRKYKEHYIGDYAQPRTR 1001
```

Db 893 MLTNLCYTYARCTKSIVPPAYAHAAAFRA-----RYTWESEMSDGGSSRSR 942

# RESULT 9

US-10-374-780A-2084  
 ; Sequence 2084, Application US/10374780A  
 ; Publication No. US20040019927A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, Bradley K  
 ; APPLICANT: Riechmann, Jose Luis  
 ; APPLICANT: Jiang, Cai-Zhong  
 ; APPLICANT: Heard, Jacqueline E  
 ; APPLICANT: Haake, Volker  
 ; APPLICANT: Creelman, Robert A  
 ; APPLICANT: Ratcliffe, Oliver  
 ; APPLICANT: Adam, Luc J  
 ; APPLICANT: Reuber, T. Lynne  
 ; APPLICANT: Keddle, James  
 ; APPLICANT: Broun, Pierre E  
 ; APPLICANT: Pilgrim, Marsha L  
 ; APPLICANT: Dubell III, Arnold T  
 ; APPLICANT: Pineda, Omaira  
 ; APPLICANT: Yu, Guo-Liang  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS  
 ; FILE REFERENCE: MBI-0047 CIP  
 ; CURRENT APPLICATION NUMBER: US/10/374,780A  
 ; CURRENT FILING DATE: 2003-02-25  
 ; PRIOR APPLICATION NUMBER: 09/837,944  
 ; PRIOR FILING DATE: 2001-04-18  
 ; PRIOR APPLICATION NUMBER: 60/310,847  
 ; PRIOR FILING DATE: 2001-08-09  
 ; PRIOR APPLICATION NUMBER: 09/934,455  
 ; PRIOR FILING DATE: 2001-08-22  
 ; PRIOR APPLICATION NUMBER: 60/336,049  
 ; PRIOR FILING DATE: 2001-11-19  
 ; PRIOR APPLICATION NUMBER: 60/338,692  
 ; PRIOR FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: 10/171,468  
 ; PRIOR FILING DATE: 2002-06-14  
 ; PRIOR APPLICATION NUMBER: 10/225,066  
 ; PRIOR FILING DATE: 2002-08-09  
 ; PRIOR APPLICATION NUMBER: 10/225,067  
 ; PRIOR FILING DATE: 2002-08-09  
 ; PRIOR APPLICATION NUMBER: 10/225,068  
 ; PRIOR FILING DATE: 2002-08-09  
 ; NUMBER OF SEQ ID NOS: 2906  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 2084  
 ; LENGTH: 969  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; FEATURES:  
 ; OTHER INFORMATION: G1149 Paralogous to G1146  
 ; US-10-374-780A-2084

Query Match 10.8%; Score 575.5; DB 4; Length 969;  
 Best Local Similarity 23.2%; Pred. No. 2.5e-34;  
 Matches 235; Conservative 161; Mismatches 339; Indels 279; Gaps 40;  
 QY 38 KKVLLVNWPKFSKSKLYDREYIEYEVMTKEVLRKRGKPPKTEIPIPDRAKLFWQHL 97  
 Db 158 KKVVRAN--HFLVQVADRDLHYDVSINPEVTSK----- 190  
 QY 98 RHEKKQTDFTLEDYVDEKDTVYSVCKRLNTVTSKMLVSEKVKVKKDSE---KKDEKLEKK 154  
 Db 191 -----TVNRNVKLLVKN---YKDSHLGSKSPAYD-GRK 220  
 QY 155 ILVTM--ILTYRKKFHLNFSRENPEKDEANRSYKF-LKNV-----MTQKVRVA 200  
 Db 221 SLYTAGPLPDSKEFVNVLAEKRADSGSKDRPFKVAVKNTSTDLVQLQQLDRKQREA 280  
 QY 201 PFVNEIKVQFAKNFYVDNNLSILRVPESHDPNRFQSL-----EVAPRIEAWFGI 251

Db 281 FY--DTIQVLDDVLRDRKPSNDYVSVGRSP-----FHTSLGKARDRGELGDGIEYWRGY 333  
 QY 252 YIGIKELFDGEPVLNPAIFVDKLFYNAPKMSLLDYLILLVDPSCNDVDVRLDKTKLMAGK 311  
 Db 334 QSLR-LTQMGLSLNIDVSARSFY-----EFIVVTDIFSKFLNIRDL--N 375  
 QY 312 MTIRQAARPRIRQLLENLKLKCAEVWDNEMSLRTERHLTFLDLCEENSLVYKVTGKSD-- 369  
 Db 376 RPLRDSRLKVKVLRTRTKVLLH-WNGTKSA-----KISGISLPL 415  
 QY 370 ----RGRNAKYYDTTLFKIYEENKCP-IEPPLPLVVKVSKAKAYAVPMHEHLEHVEKQOR 424  
 Db 416 IRELFTLEDKSEKTVVQYFAEKYVRKYQALPAIQTGSDTFRVPLFMELCQIDE-QOR 474  
 QY 425 YKNRIDLVMDQKELKRAIRKPHDYKENTKMLKELDFSSBELNVEFRGLCKLQMBICP 484  
 Db 475 YTKELNEKQVYALLKATCQRPDP-RENSIKVLVKNYNDLDS--KEFGHSTVTLASIE 531  
 QY 485 GKVLKEPMLVSNVEQIKMTPTVIRGFOBKQLNVVPEKELCCAVFVNNETAGNPCLEENDV 544  
 Db 532 ARVLPPEMLKYHDSGKEKMNPRLG---QWNMDKK----- 564  
 QY 545 VKFYTELIGCKFRGIRIGANENRGAQSIIMYDATKNEYAFYKNTLNTGTGRPIATAEA 604  
 Db 565 ----MVNGAK--VTSWTCBFKQPAIPFISCPPEH-----IEALLD- 600  
 QY 605 KMMFERLPDKEQKVLAFIISKQLNAYGFVGHVCDHTIGVANOHITSETVTKALASLRH 664  
 Db 601 --IHKAPGLQ----LLIVLPDVTGSYGKIKKICETELGIVSQCCQPRQVNL----- 648  
 QY 665 EKGSKRIFVQIALKINAKLGGINQELDWSIABIPEEKERRKTMPL-----TMYVGIDV 719  
 Db 649 ---NKQVMENVALKINVKTGGRNTVLN-----DAIRNIPLIIDRPTIIMGADV 694  
 QY 720 THPTSYSGIDYSTAAVVASIN--PGCTIYRNWIVTQ---EZ-----CRPGERAVAHGR 767  
 Db 695 THPQGEDSSPSIAAVVASMDWPEINKRYGLVSAQAHRREIIQDLYKLVDPQQRGLVH-- 752  
 QY 768 ERTDILEAKFVKLLREFAEENNDRAPAHVYVYDGVSDSEMLRVSHDELRSLSKSEVKQFM 827  
 Db 753 --SGLIREHPIAFRA-----TCQIPQRIIFYRDGVSEGFQSVLLHEMTAIRKACNSLQ 805  
 QY 828 SERDGEDPEPKYTFIVIQKRNHTRLLRRMEKDKPVVNVKOLTPAETDVAAVAKQWSEDMK 887  
 Db 806 ----ENYVPRVTFVIVQKRNHTRLLRRMEKDKPVVNVKOLTPAETDVAAVAKQWSEDMK 836  
 QY 888 ESKETGIVAPSSGTTVDKLVSKYKDFDFFLASHHVLGTSRPGHYTMYDDKMGMSQDEVY 947  
 Db 837 --DKSGNIQP--GTVVDTKICHNEFDFYLNSHAGIQGTSRPAHYHVLDDENGFTADQLQ 892  
 QY 948 KMTYGLAFLSARCKRPISLPVPVHYAHLSCAKELVRYTKEHYIGDYAQPRT 1001  
 Db 893 MLTNLCYTYARCTKSIVPPAYAHAAAFRA-----RYTWESEMSDGGSSRSR 942

RESULT 10  
 US-10-412-699B-752  
 ; Sequence 752, Application US/10412699B  
 ; Publication No. US20040045049A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mendel Biotechnology, Inc.  
 ; APPLICANT: Zhang, James  
 ; APPLICANT: Fromm, Michael E.  
 ; APPLICANT: Heard, Jacqueline E.  
 ; APPLICANT: Riechmann, Jose Luis  
 ; APPLICANT: Adam, Luc J.  
 ; APPLICANT: Broun, Pierre E.  
 ; APPLICANT: Pineda, Omaira  
 ; APPLICANT: Reuber, T. Lynne  
 ; APPLICANT: Keddle, James S.  
 ; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Jiang, Cai-Zhong

```

; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creelman, Robert A.
; APPLICANT: DuBell, Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumimoto, Roderick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: polynucleotides and polypeptides in plants
; FILE REFERENCE: MBI-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B
; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 752
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1149
; US-10-412-699B-752

Query Match          10.8%; Score 575.5; DB 4; Length 969;
Best Local Similarity 23.2%; Pred. No. 2.5e-34;
Matches 235; Conservative 161; Mismatches 339; Indels 279; Gaps 40;

QY 38 KVVLLVNNFKSSKIYDEYEVYKTVKVLNRPKPKPKPKTEIPIDRAKLFQWHL 97
DB 158 KKVVRAN--HELQVADRDLHYDVSINPEVSK----- 190
QY 98 RHEKQTDFILEDYVDEKDTVYVSVCLNTVTSKMLVSEKVVKKDSE---KKDEKDKLEK 154
DB 191 -----TVNRNVNKKLLVKN---YKSHLGKGPAYD-GRK 220
QY 155 ILYTM--ILYTKKFKHLNFSRENPERDEANRSYKF-LKNV-----MTQKVRYA 200
DB 221 SLYTAGPLPDSKEFVNNLAERADSGSKDRPFKVAKNVSTLDYLQQLQFLDKRQEA 280
QY 201 PFVNEIKVQAKNFYVNNLSILRVPSFHDNRFQSL-----EVAPRIEAWFGI 251
DB 281 PY--DTIQVLDVLRDKPSNDYVSVGRSF-----FHTSLGKDARDGELGDGIEYWRGY 333
QY 252 YIGIKELFDPGEPLNPAIVDKLFPYNAPKMSLLDYLILLIVDQSCNDNDVRKDKTKLMA 311
DB 334 FOSLR-LTQWGLSLNTDVSARSFY-----EPIVVTDFISKFLNIRDL--N 375
QY 312 MTRQARPRIROLLENKLKAEVWMDNEMSLRHLRFLDLICEENSLVYKVTGKSD-- 369
DB 376 RPLRDSRLKVKVKKLRTLVKLLH-WNGTSA-----KISGSSLP 415
QY 370 ----RGRNAKVDITLTKIYENKPF-IEFPHPLPVKVKSGAKYAVPWEHLEVHEKPOR 424
DB 416 IRELPTLEDKSEKTVVQVFAEKYINRVYQALPAITQGTSDTRPVVLPMLCQIDR-GQR 474

```

```

QY 425 YKNRIDLVNQDKFLKRAIRKPHDYKENTIKMLKELDFSEELNVERFGLCSKLOMIECP 484
DB 475 YTKRLNEKQVTLAKATCORPPD-RENSIGNLVNKNYNDLDS---KEFGNSVTTQLASTE 531
QY 485 GKVLKEPMLVNSVNEQIKMTPTVIRGFOBKQLNVNPEKELCCAVFVVVNETAGNPLCEENDV 544
DB 532 ARVLPPMLKYHDSGKEMVNPRLG---QNMWIDKK----- 564
QY 545 VKFYTELIGGCKPGRGIRIGANENRGAQSIWYDATKNYAFYKNCNTLNTGIGRFEIAATEA 604
DB 565 -----MVNGAK---VTSWTCFKPQPAIPFISCPPEH-----IEEALLD- 600
QY 605 KUMFERLPDKEQKVLMEFIISKRQLNAYGFVGHYCDHTIGVANQHIITSETVTVKALASLRH 664
DB 601 --IHKRAFGQ-----LLVILPDVTGSGYKIKRI CETELGIVSQCCQPRQVKNL----- 648
QY 665 EKGSKRIFYQIALKINAKLGGINBELDMSEIAEISPEEKERKTWPL-----TMYVGIDV 719
DB 649 ---NKQYMENVALKINVKTGGRNTVLN-----DAIRNPIPLITDRPTIIMGADV 694
QY 720 THPTSYSGIDYSIAAVASIN--PGGTIYRNMIVTQ---EE-----CRPGERAVAHGR 767
DB 695 THPQGEDSSPSIAAVASMDWPEINRYKRGVLSAQAHREEIIODLYKLVQDPQRLVH-- 752
QY 768 ERTDILEAKFVKLLREFAENNDRAPAHIVVVRDGVSDSEMLRVSHDELRLSKSEVKQFM 827
DB 753 --SGLIREHFIAPRA-----TGQIPQRIIFRDGVSEGFQSOVLLHENTAIRKACNSLQ 805
QY 828 SERDGEDPEPKYFTIVIOQRNTRLRLRMEKDKPVVKNDLTPAETDVAVAVKQWDEDMK 887
DB 806 ----ENVVPRVTFVIVQKRHHTRLFPPEQHG-----NRDMT----- 836
QY 888 ESKETGVNPSGTTVDKLIIVSKYKFDPFSLASHHGVLTGTSRPGHYTVMYVDDKMSODEVY 947
DB 837 --DKSGNIQP--GTVDYTKICHNPBDFYLNSHAGIQGTSRPAHYHVLVDENGFTADQLQ 892
QY 948 KMTYGLAFLSARCRKPISLPVVPHVAHLSCEKAKELYRTRYKEHYIGDYAQPRTR 1001
DB 893 MLTNLCTYARCTKSVSIVPAYVAHLAAPRA----RYMESEMSDGGSSRSR 942

RESULT 11
US-10-467-397-7
; Sequence 7, Application US/10467397
; Publication No. US20040137448A1
; GENERAL INFORMATION:
; APPLICANT: THORNTON, Michael; HAFALIA, April J. A.;
; APPLICANT: LU, Dying Aina M.; ARVIZU, Chandra S.;
; APPLICANT: SWARNAKAR, Anita; LU, Yan;
; APPLICANT: WARREN, Bridget A.; BAUGHN, Maria R.;
; APPLICANT: TANG, Y. Tom; LEE, Ernestine A.;
; APPLICANT: YAO, Monique G.; RAMKUMAR, Jayalaxmi;
; APPLICANT: KHAN, Farrah A.; GANDHI, Ameena R.;
; APPLICANT: DING, Li; YUE, Henry;
; APPLICANT: GIETZEN, Kimberly J.; CHAWLA, Narinder K.;
; APPLICANT: THANGAVELU, Kavitha; ELLIOTT, Vicki S.
; APPLICANT: MARQUIS, Joseph P.
; TITLE OF INVENTION: NUCLEIC ACID-ASSOCIATED PROTEINS
; FILE REFERENCE: PI-0368 USN
; CURRENT APPLICATION NUMBER: US/10/467,397
; CURRENT FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: PCT/US02/03844
; PRIOR FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US 60/268,118
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/270,963
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/270,858
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/271,194
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/274,071

```







```

QY 795 HIWVRDGVSDSEMLRVSHDELSRLKSEVKQPMSESDGEDPZ--PKYTFIVIOKRHNTKL 852
DB 798 RIILYRDDGVSEGQFPFVLQHELTAREACIKL-----EPEYRPGCITFIVVQKRHNTKL 840
QY 853 LRRMEKDKPVVWVKDLTPAETDVAAAVRQWBEEDMKESKETGVNPFSSGTTVDKLIVSKYK 912
DB 841 FCAEKE-----QSGKSG--NIPAGTIVDVGIHTPTE 870
QY 913 FDFFLASHHGVLGTSRPGHYTMYDDKGMSEVYKMTYGLAFLSARCKPISLPPVPHY 972
DB 871 FDFYLCSHQGIGQTSRPSGHYHVLWDDNHPDSDELQCLTVQLCHTYVRCRVSVIPAPAY 930
QY 973 AHLSCSEKAKELYRTYKEHYIGD 994
DB 931 AHLVAFRAR-YHLVEKEHDSGE 951

RESULT 14
US-11-097-143-27927
; Sequence 27927, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27927
; LENGTH: 984
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-27927

```

RESULT 15  
US-10-645-746-9  
; Sequence 9, Application US/10645746  
; Publication No. US20040265839A1  
; GENERAL INFORMATION:  
; APPLICANT: Mello, Craig C.  
; APPLICANT: Tabara, Hiroaki  
; APPLICANT: Grishok, Alla  
; APPLICANT: Fire, Andrew  
; TITLE OF INVENTION: RNA INTERFERENCE PATHWAY GENES AS TOOLS FOR TARGETED GENETIC  
; TITLE OF INVENTION: INTERFERENCE  
; FILE REFERENCE: UMY-052DV1  
; CURRENT APPLICATION NUMBER: US/10/645,746  
; CURRENT FILING DATE: 2003-08-20  
; PRIOR APPLICATION NUMBER: US 09/689,992  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 60/193,218  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: US 60/159,776



```
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-645-746-9

Query Match      10.4%; Score 557; DB 5; Length 766;
Best Local Similarity 23.4%; Pred. No. 4.5e-33;
Matches 184; Conservative 139; Mismatches 312; Indels 150; Gaps 23;

QY 230 HDNPEQSLVAPRIEAWFGIYIGIKELFDGEPVNFVDFKLFYNAPKMSLLDYLLLI 289
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
89 HSAGQTHAESKLGGRGVWFGHQSVRP--SQWMLNIDVSAFVRS--MPVIEFIAEV 145
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
290 VDPQSCNDVDRKDKTKLMAGKMTIRQAARPRIROLLENLKLKCAEVMDNEMSLRTERHL 349
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
146 L-----ELPQALAEARRALSDAQRVKFTKEIRGLKIEITH----- 180
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
350 TFLDLCEENSLVYKTKSGDRGNKKY-----DTTLFK-IYEENKKPIBFPHP 398
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
181 -----CGQMRKRYVNCVTRRPAQTQTFPLQLETGTQIECTVAKYFYDKYRIQLKYPHLP 235
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
399 LVKVKSGAKEYAVPMHEHLEVEKHPQRYKNRIDLVMDQKELKRAIKRPHDYKENTLMLKE 458
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
236 CLOVGOEQKHLYLPPEVCNI-VFGORCIKKLTDVOTSTWIKATARSAPEREREISNLVRK 294
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
459 LDFSSBELNFVERFGLCSKLOMECPGKVLKEPMLVNSVNEQIKWTP-----VIRGFQEK 513
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
295 AEFSDAP--FAHEFGITINPANTEVKGRVLSAPKLLYGGRTATALPNQGVWDMRG---K 349
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
514 QLVNVPEKELCCAVFVNETAGNCPLENDVVVKFYTELIGGCKFRGIRIGANENRGAOSI 573
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
350 QPHT---GIDVRVWAIACFAQQOHVKENDLRFMTNQL-----QRI 386
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
574 MYDATKNEVAFYKNCNTLNTGIRPEIAATEAKNMPERLPDKEQKVLMTFIIISKROLNAYG 633
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
387 SNDAGMPIVG-----NPFCKYAVGEVQEPMPFKYLKQNYSGIOLVVVVLPGKTPVYA 439
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
634 FVKHYCDHTIGVANQHIITSETVTKALASLRHEKSGKRIFYQIALKINAKLGGINGELWS 693
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
440 EVKRVGDTVLGIATQCVQAKNAIRT-----TPQTLNLCLENNVVLGGVNSIL--- 487
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
694 ETAEISPEEKERRKTMPLTMVYGVIDVTHPTSYSGIDYSIAAVASINPGGTIYRNMIVTQ 753
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
488 -----LPNVRPRIFNEPV-IPFGCDITHPPAGDSRKPSTAAVVGSMDAHPSRYAATVRVQ 541
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
754 EECRGERAVAGRETRDILEAKFVKLLREFAENNDNRAPAHIVVYRDGVSDSEMLRVSH 813
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
542 Q-----HRQELISDLTYMWRELLVQFYRTRFRK--PARIVYRDGVSEGQFFNVLQ 590
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
814 DELSLKSEVKQFMSERGEDPEPKYTFIVIOKRHNTRLRRMEKDKPVVKNKDLTPAETD 873
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
591 YELRAIREAC--WMLER---GYQPGITTFIAVQKRHHTRLFAVDKKDQ--VGK----- 635
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
874 VAAVAVKQWEDMKESKETGI VNPSSGTTVDKLI VSKYKFPDFLASHHGVLGTSRPHYT 933
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
636 -----AYNIPP GTTVDVGITHTPTEFDYLCSHAGIQGTSRPSHYH 675
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
934 VMYDDKGMQSDEVYKMTYCLAFSLARCKPISLPVPVHYAHLSCAKELRYTYKEHYIG 993
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
676 VLWDDNNLTADLEQLLTQMCHTYVRCRNSVSIAPAPYYAHLVAFRAR--YHLVDREHDSG 734
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
994 DYAAQP 998
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
735 EGSQP 739
```

Search completed: July 5, 2006, 13:49:40  
Job time : 195 secs



source

```

1. 3229
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/db_xref="taxon:6239"
/chromosome="V"
/clone="yk296b10"
1. 3229
/gene="rde-1"
21..3083
CDS
/gene="rde-1"
/note="similar to Caenorhabditis elegans K08H10.7"
/codon_start=1
/product="RNA interference promoting factor RDE-1"
/protein_id="AAF06159.1"
/db_xref="GI:6272678"
/translation="MSSNPPELEKGFYRHSLDPEMKWLARPTGKDGKPYEKKVLLV
NPFSSKIDYEYIEYVKMTKEVLNRRKPGFPFKTBIPIPDRAKLPWQHLREKK
QTFDLELVFDEKOTVSVCLNTVTSMKVSEKVKVKKSEKKDEKLEKILVTMI
LTYRKFLHNSRENPEDDEANRSYKFLKNVTKVYAPVFNSEIKVQPAKNFVYD
NNSILVRPESFDHPNRPESQLEVAPRIENWFGIYIGIKELFDGEVPLNPAIVDLFYN
APMSLDYLLLI VQSCNDVDRKDLTKLMAGKMTINQAAPRIQLLENLKKCA
EYWDNMSRLTHERHUTFDLCENSLVYKVGKSDRGNAKYDTFLFKIYBENKKFI
EFPFLPVKVSQGAKEYAVPMHELVHEKQRYQRIIDLVMQDKFLKRAIRKPHDYKE
NTLKMELDFSEELNFERFGLCSKLQIECPGVKLEPMLVNSVEQIKNTVPIR
GFOEQNLNVPSKELCCAVFVNVTAGNPCLNEVDVVFYTELIGCKPFRIGRANE
NRGAQSIYDATKNEYAFYKNCITLNTGIRFEIATKAEAKNMEPERLPDKEQKLMFII
SRQNLNAYGVFKYGHDTIGVANQHTSETVTKALASLRHKGSRIFQIYALKINAK
LGGINQELDWSIAIEISPEEKRRKTMPLTVYGDVTHPTSISGIDYSIAAVASIN
PGGTIYRNMI VQSECRPGERAVAHGERTDILEAFVKFLRFEPAENDRNPAHIVV
YRDGYSMLVSHDESLRSLKSEKQFMSERDGPBPKYTFI VIQKHNTRLRRM
EKDKPVNKDLTPAETDVAAVAKOWEEDKESKETI VNPSSGTTVDKLI VSKYKFD
PFLASHGVLTGTSRPHGVTVYDDKMSODEVYKNTYGLAFLSARCKPISLPVPVHY
AHLSEKAKELRTYKEHYIGDYAOPRTHMEHFLQTNVKYPGMSFA"

```

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	3229
Score:	5349.00	Matches:	1020
Percent Similarity:	100.0%	Conservatives:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	13	Gaps:	0

US-10-645-746-3 (1-1020) x AFI80730 (1-3229)

QY	1	MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspProGlu	20
DB	21	ATGTCCTCGAAATTTCCCGAAATTGGAAAAAGGATTTTATCGTCAITTCCTCGATCCGGAG	80
QY	21	MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysVal	40
DB	81	ATGAAATGGCTTCGAGGGCCCACTGTGTAATCGACGGCAAAATTTCTATGAGAAGAAGTA	140
QY	41	LeuLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrTyrGlu	60
DB	141	CTTCTTTTGGTAAATTTGGTTCAGTTCTCCAGCAAAATTTACGATCGGGAATACTACGAG	200
QY	61	TyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLys	80
DB	201	TATGAAGTCAAAATGACAAAGGAAGTATTGAATAGAAAAACCAGGAAAAACCTTTCCCAAAA	260
QY	81	LysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGlu	100
DB	261	AAGACAGAAATTCCAATTCCTGATCGTGCAAAACCTCTTCTGGCAACATCTTCGGCATGAG	320
QY	101	LysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAspThrValTyr	120
DB	321	AGAAGCACACAGATTTTATCTCGAAGACTATGTTTTTTGATGAAAGGACACTGTTTAT	380
QY	121	SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysValValLys	140
DB	381	AGTGTTTGTTCGACTGAACACTGTGCATCAATCAAAAATGCTGGTTTCGGAGAAGTAGTAAAA	440

QY	141	LyAspSerGluLysLysAspGluLysAspLeuGluLysLysIleLeuTyrThrMetIle	160
DB	441	AGGATTTGGAGAAAAAAGATGAAAAGGATTTGGAGAAAAAATCTTATACACAAATGATA	500
QY	161	LeuThrTyrArgLysPheHisLeuAsnPheSerArgGluAsnProGluLysAspGlu	180
DB	501	CTTACCTATCGTAAAAAAATTTCCCTGAACCTTTAGTCGAGAAAAATCCCGAAAAAGACGAA	560
QY	181	GluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysValArgTyrAla	200
DB	561	GAAGCGAATCGGAGTTTACAAATTCCTGAAGAAATGTTATGACCCAGAAAGTTTCGTACGGC	620
QY	201	ProPheValAsnGluLysValGlnPheAlaLysAsnPheValTyrAspAsnAsn	220
DB	621	CTTTTGTGAACGAGAGATTAAGTACAAATTCGCGAAAAAATTTTGTGTACGATAATAT	680
QY	221	SerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGlnSerLeuGlu	240
DB	681	TCAATTTCTCGAGTTCTCTGAATCGTTTACGATCCAAACAGATTCGAACAATCATTAGAA	740
QY	241	ValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyLysGluLeuPheAsp	260
DB	741	GTAGCACCAGAAATCGAAGCATGGTTTGGAAATTTACATTTGGAAATCAAGAAATTTGTTTCGAT	800
QY	261	GlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAlaProLysMet	280
DB	801	GGTGAACCTGTGCTCAATTTTGCATTTGTCTGAATAACTATTCTCAATGTCACCGAAAAATG	860
QY	281	SerLeuLeuAspTyrLeuLeuLeuValAspProGlnSerCysAsnAspAspValArg	300
DB	861	TCTCTCTCGATTATCTTCTCTTAATTTGTGCGACCCCGCATGCTGTAAACGATGATGACGA	920
QY	301	LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgPro	320
DB	921	AAAGATCTTAAACAAAACTGATGCGGGAAAAATGACAAATCAGACACCGCGCGGCCA	980
QY	321	ArgIleArgGlnLeuLeuLeuLysLysCysAlaGluValTrpAspAsnGlu	340
DB	981	AGNATTCGACAAATTTATTTGGNAAATTTGAAGCTGAAATGCGCAGAGTTTGGGATACGAA	1040
QY	341	MetSerArgLeuThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeu	360
DB	1041	ATGTCGAGATTGACAGAACGACATCTGACATTTCTAGATTTTGTGCGAGGAAAACTCTCT	1100
QY	361	ValTyrLysValThrGlyLysSerAspArgGlyArgAsnAlaLysIleTyrThrPheThr	380
DB	1101	GTTTATAAGTCACCTGGTAAATCGGACAGAGGAGAAATGCAAAAAGTAGTACGATCTACA	1160
QY	381	LeuPheLysIleTyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuVal	400
DB	1161	TTGTTTCAAAATCTATGAGGAAAAACAAAAGTTTCATTTAGTTTCCCCACCTACCATAGTC	1220
QY	401	LysValLysSerGlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGlu	420
DB	1221	AAAAGTTAAAGTGGAGCAAAAGATACGCTGTACCAATGGAACATCTTGAAGTTCTAGAG	1280
QY	421	LysProGlnArgTyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArg	440
DB	1281	AAGCCACAAAAGATACAAGAAATCGAATTTGATCTGGTGATCGAAGACAAAGTTTCTAAAGCGGA	1340
QY	441	AlaThrArgLysProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAsp	460
DB	1341	GCTACACGAAACCTCAGACTACAAAGAAAAATACCTTAAAAATGCTGAAGAATTTGGAT	1400
QY	461	PheSerSerGluGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMet	480
DB	1401	TTCTCTCTCGAAGAGCTAAATTTTGTGTGAAGATTTGGATTTATGCTCCAAACTTCAGATG	1460
QY	481	IleGluCysProGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGln	500
DB	1461	ATCGAATGTCCAGGAAGGTTTGTGAAGAGCCAAATGCTTGTGANTAGTGTAAATGAACAA	1520
QY	501	IleLysMetThrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGlu	520

```

Db 1521 ATTTAAATGACACCAAGTATTCGTGGATTTCCAGAAAACAAATGAAATGCTGTTCCCGAA 1580
Qy 521 LysGluLeuCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGlu 540
Db 1581 AAAGAACCTTGTCTGCTGCTGTTTGTAGTCAACGAAACAGCGGGAATCCATGCTTAGAA 1640
Qy 541 GluAsnAspValValLysPheThrGluLeuLeuGlyCysLysPheArgGlyLe 560
Db 1641 GAGAACGACGTTGTTTAAGTCTCACCCGAACTAATGGTGTTGCCAAGTTCCTCGGAATA 1700
Qy 561 ArgileGlyAlaAsnGluAsnArgGlyValaGlnSerileMetTyrAspAlaThrLysAsn 580
Db 1701 CGAATTGGTCCCAATGAAAACAGAGGAGCGCAATCTATTATGATACGACGCGCAAAAAT 1760
Qy 581 GluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyLeGlyArgPheGluLeAla 600
Db 1761 GAATATGCTCTTACAAAAATTTGACACATAATACCGGAATCGGTAGATTGGAATAGCC 1820
Qy 601 AlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLysValLeuMet 620
Db 1821 GCACAGAGGAGGAAGATATGTTTGAACGCTTCCCGATAAAGAACAAAAAGTCTTAATG 1880
Qy 621 PheleileileSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHisTyrCysAsp 640
Db 1881 TTCATTATCATTTCCAAACGCACTGAAATGCTTACCGTTTGTGGAACCAATTATTCGAT 1940
Qy 641 HisThrileGlyValAlaAsnGlnHisileThrSerGluThrValThrLysAlaLeuAla 660
Db 1941 CACACCATCGGTAGCTAATCAGCATATTACTTCTGAAACAGTCAACAAAGCTTTGGCA 2000
Qy 661 SerleuArgHisGluLysGlySerLysArgilePheTyrGlnleAlaLeuLysIleAsn 680
Db 2001 TCACTAAGGCACGAGAAAGATCAAAACGAAATTTCTATCAATTTGCAATGCAATCAAC 2060
Qy 681 AlaLysLeuGlyGlyLysAsnGlnLeuAspTyrSerGluLeAlaGluLysSerPro 700
Db 2061 GCGAATTAGAGGTATTACACGAGGCTTGACTGTGTCAGAAATGCGAATAATATCACCA 2120
Qy 701 GluGluLysGluArgArgLysThrMetProLeuThrMetTyrValGlyLysAspValThr 720
Db 2121 GAAGAAAAAGAAAGACGGAACAAATGCCATTAACTATGTATGTGTTGAAATGATGTAAC 2180
Qy 721 HisProThrSerTyrSerGlyLysAspTyrSerileAlaValValAlaSerileAsn 740
Db 2181 CATCCAACTCCCTACAGTGGAAATTTGATTTCTATAGCGGCTGTAGTAGCGAGTATCAAT 2240
Qy 741 ProGlyGlyThrileTyrArgAsnMetIleValThrGlnGluCysArgProGlyGlu 760
Db 2241 CCAGTGGAACTATCTATCGAATATGATTGTGACTCAAGAAATGTCTGCCCGGTGAG 2300
Qy 761 ArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAlaLysPheValLysLeu 780
Db 2301 CGTGCACTGCTCATCGACGGGAAACAGACAGATATTTTGGACCAAGTTCTGTAATTCG 2360
Qy 781 LeuArgGluPheAlaGluAsnAsnAspArgAlaProAlaHisileValValTyrArg 800
Db 2361 CTCAGAGAAATTCAGAGAAACACAGCAATCGAGCACCGAGCATATTGTAGTCTATCGA 2420
Qy 801 AspGlyValSerAspSerGluMetLeuArgValSerHisAspGluLeuArgSerLeuLys 820
Db 2421 GACGGATTAGCGATTCGGAGATGCTACGTTGTAGTCAATGATGAGCTTCGATCTTTAAAA 2480
Qy 821 SerGluValLysGlnPheMetSerGluArgAspGlyGluAspProGluProLysTyrThr 840
Db 2481 AGCGAAGTAAACAAATTCATGTCGGAACGGATGGAGAGATCCAGACGCGAGTACAGC 2540
Qy 841 PheileValileGlnLysArgHisAsnThrArgLeuLeuArgMetGluLysAspLys 860
Db 2541 TTCATTGTGATTCCAGAAAAACACAAATACAGCATTCGTTCTCGAAGATGCAAAAGATAAG 2600
Qy 861 ProValValAsnLysAspLeuThrProAlaGluThrAspValAlaValAlaValLys 880

```

```

Db 2601 CCAGTGGTCAATAAAGATCTTACTCTCGCTGCAAAACAGATGCGTGTGCTGCTGTTAAA 2660
Qy 881 GlnTrpGluAspMetLysGluSerLysGluThrGlyLysValAsnProSerSerGly 900
Db 2661 CAATGGGAGGAGGATATGAAAGAAAGCAAACTGGAATTGTGAACCCATCATCCGGA 2720
Qy 901 ThrThrValAspLysLeuLysSerLysTyrLysPheAspPheLeuAlaSerHis 920
Db 2721 ACACTGTGGAATAAACTTATCGTTTCGAAATACAAATTCGAATTTCTTCTGTCATCTCAT 2780
Qy 921 HisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMetTyrAspAspLysGly 940
Db 2781 CATGGTCTCTTGTGTACATCTCGTCCAGGACATTACACTGTATGTATGATGAGATAAGGA 2840
Qy 941 MetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPheLeuSerAlaArgCys 960
Db 2841 ATGAGCCCAAGATGAAGTCTATAAATGACCTACGAGCTTGTCTTCTCTGCTAGATGT 2900
Qy 961 ArgLysProLysSerLeuProValProValHisTyrAlaHisLeuSerCysGluLysAla 980
Db 2901 CGAAAAACCCATCTCGTTGCTGCTTCCGGTTCAATTATGCTCATTTATCATGTGAAAAAGCG 2960
Qy 981 LysGluLeuTyrArgThrTyrLysGluHisTyrileGlyAspTyrAlaGlnProArgThr 1000
Db 2961 AAAGAGCTTATCGAATTAAGAAACATTTACATCGGTGACTATGACAGCCACGGACT 3020
Qy 1001 ArgHisGluMetGluHisPheLeuGlnThrAsnValLysTyrProGlyMetSerPheAla 1020
Db 3021 CGACAGAAATGGAACATTTTCTCCAACTTAACGTGAAGTACCTCGAATGTCTGTTCCGA 3080

```

## RESULT 2

CEK08H10/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

CONSTRM

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CEK08H10 39339 bp DNA linear INV 09-AUG-2005  
Caenorhabditis elegans Cosmid K08H10, complete sequence.

283113  
283113.1 GI:3217648  
HTG.

Caenorhabditis elegans  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 39339)

C. elegans Sequencing Consortium  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology

Science 282 (5396), 2012-2018 (1998)

9851916  
2 (bases 1 to 39339)  
Gardner,A.E.

Direct Submission  
Submitted (27-NOV-1996) Nematode Sequencing Project, Sanger  
Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
worm@sanger.ac.uk

On Jun 13, 1998 this sequence version replaced gi:1695070.  
Coding sequences below are predicted from computer analysis, using  
predictions from Genefinder (P. Green, U. Washington), and other  
available information.

Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.

For a graphical representation of this sequence and its analysis  
see: [http://www.wormbase.org/perl/ace/elegans/seq/sequence?](http://www.wormbase.org/perl/ace/elegans/seq/sequence?name=K08H10;class=Sequence)  
name=K08H10;class=Sequence

This sequence is the entire insert of clone K08H10. The true left  
end of clone T0688 is at 32990 in this sequence. The true right end



TTYLSLRDGVOPEDNOPLVLAKKIKGMEVSWGGLAFATANEPOLTEGAIKTALSSSE  
 PNSTVVDQIEDLYSGTNQVLSFNIIISV"  
 join(24915..24953,26257..26580)  
 /locus\_tag="K08H10.3"  
 join(24915..24953,26257..26580)  
 /locus\_tag="K08H10.3"  
 /standard\_name="K08H10.3"  
 /notes="contains similarity to Pteronotus parnellii Brcal  
 (Fragment).; TR:Q7YQ11"  
 /codon\_start=1  
 /product="Hypothetical protein K08H10.3"  
 /protein\_id="CAB05542.1"  
 /db\_xref="GI:3924818"

## Alignment Scores:

Pred. No.: 0 Length: 39339  
 Score: 5001.00 Matches: 1016  
 Percent Similarity: 85.2% Conservative: 3  
 Best Local Similarity: 85.1% Mismatches: 3  
 Query Match: 93.5% Indels: 177  
 DB: 13 Gaps: 10

US-10-645-746-3 (1-1020) x CEK08H10 (1-39339)

QY 1 MetSerSerAsnPheProGluLeuGluLysGlyPheTyrArgHisSerLeuAspPro--- 19  
 DB 4606 ATGTCCTCGAATTTTCCGGAATTGGAAAGGATTTTATCGTCATTCTCGATCCGGTA 4547  
 QY 20 -----GluMetLys 22  
 DB 4546 TGATCAATTATTAGCAGCTATAGATATATAAGTTTGATATTATATTATAGAGATGAA 4487  
 QY 22 sTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGluLysLysValLeuLe 42  
 DB 4486 ATGGCTTGGAGGCCCACTGGTAAATCGCAGCGCAAAATTCATAGAGAAAGTACTTCT 4427  
 QY 42 uLeuValAsnTrpPheLysPheSerSerLysLysLeuTyrAspArgGluTyrGluTyrG 62  
 DB 4426 TTTCGTAATTTGGTTCAAGTTCTCCAGCAAAATTTTACGATCGGGAATCTACGAGTATGA 4367  
 QY 62 uValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysTh 82  
 DB 4366 AGTGAATAATGACAAAGAGATTTGAATAGAAACCCAGGAAACCTTTCCCAAAAAGAC 4307  
 QY 82 rGluLeuProIle-----ProAs 88  
 DB 4306 AGAAATTCCAATGTAAGTGTCTGTAATTTAGTCAAACTAAATTTTATTTTTCAGTCCCGA 4247  
 QY 88 pArgAlaLysLeuPheTrpGlnHisLeuArgHisGlyLysLysGlnThrAspPheIleLe 108  
 DB 4246 TCGTGCAAAACTCTTCTGGCAACATCTTCGGCATGAGAGAGACACACAGATTTTATTTCT 4187  
 QY 108 uGluAspTyrValPheAspGluLysAspThrValTyrSerValCysArgLeuAsnThrVa 128  
 DB 4186 CGAAGACTATGTTTTGATGAAAGGACACTGTTTATAGTGTTCGACTGACACTGT 4127  
 QY 128 lThrSerLysMetLeuValSerGluLysValLysLysValLysLysAspSerGluLysLysAspG 148  
 DB 4126 CACATCAAAATGCTGGTTTCGGAGAAAGTAGTATAAAAGGATTCGGAGAAAAAGATGA 4067  
 QY 148 uLysAspLeuGluLysLysLysLeuTyrThrMetIleLeuThrTyrArgLysLysPheH 168  
 DB 4066 AAGGATTTTGGAGAAAAAATCTTATACCAATGATACCTTACTATCGTAAAAAATTTCA 4007  
 QY 168 sLeuAsnPheSerArgGluAsnProGluLysAspGluAlaAsnArgSerTyrLysPh 188  
 DB 4006 CTGGAATTTTGTGAGAAATCCGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3947  
 QY 188 eLeuLys-----AsnValMe 193  
 DB 3946 CCTGAA-GGTTTTATGAAAAACCGCATTTATACAAAACAAATTTAGCTTTTCAGAAATGTTAT 3888  
 QY 193 tThrGlnLysValArgTyrAlaProPheValAsnGluLysVal----- 209

DB 3887 GACCCAGAAAGTTGCGTACGCGCTTTTGTGAACAGGAGATTTAAAGTGTGAGTTGCAAT 3828  
 QY 210 -----GlnPheAlaLysAsnPh 215  
 DB 3827 AATAATAATAATAATCACTCACTCACTATATATATATTTTAAAGACANATTCGCGAAATTT 3768  
 QY 215 eValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPh 235  
 DB 3767 TGTGTACGATAATAATTAATTCATTTCTCGAGTTCTGTAATCGTTTCAACGATCCAAACAGATT 3708  
 QY 235 eGluGlnSerLeuGluValAlaLysProArgIleGluAlaTrpPheGlyIleTyrIleGlyI 255  
 DB 3707 CGAACCAATCATTTAGAGTAGCACCAAGATCGAAGCATGGTTTGGAAATTTACATTTGGAAT 3648  
 QY 255 eLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIle----- 269  
 DB 3647 CAAGAATTTGTCATGGTGNACCTGTGCTCAATTTTGCAAG-TAAGTTTGAGAACTGC 3589  
 QY 270 -----ValAspLysLeuPheTyrAsnAlaProLys 279  
 DB 3588 GATAAAAAATCATGTGATTTTGTGCAAGTTGTGATATAAATCTTCTACAAATGCACCGAA 3529  
 QY 279 sMetSerLeuLeuAspTyrIleLeuLeuValAspProGlnSerCysAsnAspAspVa 299  
 DB 3528 AATGTCTCTTCTGGATTATCTCTCTTAATTTGCGACCCCGCAGTCGTGTAAACGATGATGT 3469  
 QY 299 lArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaAr 319  
 DB 3468 ACGAAAGATCTTAAACCAAACTGATGGCGGAAATGACAAATCAGACAGCCGCGCG 3409  
 QY 319 gProArgIleArgGlnLeuLeuGluAsnLeuLysCysAlaGluValTrpAspAs 339  
 DB 3408 GCCAAGAAATTCGACAATTTATGGAAATTTGAAGCTGAAATGCGCAGAACTTTGGGATAA 3349  
 QY 339 nGluMet-----SerArgL 344  
 DB 3348 CGAAATTTAGTTTAAATTTATTTCAAAATAATATATACAAATTTTTCAGGTCCGAGAT 3289  
 QY 344 euThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysV 364  
 DB 3288 TGACAGAACGACATCTGACATTTCTAGATTTTGGCAGGAAACTCTCTCTTTTATTAAG 3229  
 QY 364 alThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysI 384  
 DB 3228 TCACTGGTAAATCGGACAGAGAAAGAAATGCAAAAGTAGCATCTACATTTGTTCAAAA 3169  
 QY 384 leTyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysS 404  
 DB 3168 TCTATGAGGAAAAACAAAAGTTTCATTTGAGTTTCCCACCTACCCACTAGTCAAAAGTTAAAA 3109  
 QY 404 exGlyAlaLysGluTyrAlaValProMetGluHisLeuValHisGluLysProGlnA 424  
 DB 3108 GTGGAGCAAAAGANATACGCTGATCCATGGAACATCTTGAGTTTCATGAGAGCCACAAA 3049  
 QY 424 rGlyTyrAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgL 444  
 DB 3048 GATACAGAATCGAATTTGATCTGGTGATCGAAGCAAGTTTCTTAAAGCGAGCTACACGA 2989  
 QY 444 yProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerG 464  
 DB 2988 AACCTCACGACTACAAAGAAAAATACCTTAAAAATGCTGAAGAAATTTGGATTTCTCTCTG 2929  
 QY 464 luGluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysP 484  
 DB 2928 AAGACTAAATTTTGTGGAAGATTTGGATTTATGCTCCAAACTTCAGATGATCGAATGTC 2869  
 QY 484 roGlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMet 504  
 DB 2868 CAGGAAAGGTTTGAAGAGAGCCCAATGCTTGTGAATAGTGTAAATGAACAAATTAATGA 2809  
 QY 504 hrProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGluLysGluLeu 524



Db 2808 CACCACTGATTCGTGGATTTCAAGAAAAACAATTGAATGTGGTTCCTCCGAAAAAGAACTTT 2749  
 Qy 524 yscysalavalphevalvalaenGluThrAlaGlyAsnProCysLeuGluGluAenAspY 544  
 Db 2748 GCTGTGCTGTTTGTGTAGTCAACGAAACAGCGGAAATCCATGCTTTAGAGAGAACGACG 2689  
 Qy 544 alVal-----LysPheTyrThrGlu 550  
 Db 2688 TTGTGTAAAGTGTTCCTACGTAGATTATTCGGAATATATTTTCAGTAAAGTCTTACCGAA 2629  
 Qy 551 LeuileGlyCysLysPheArgGlyileArgileGlyAlaAsnGluAenArgGlyAla 570  
 Db 2628 CTAATTTGGTGGTTGCAAGTTCGTGGAATACGAATTTGGTGCCTGCAATGAAACAGAGAGCG 2569  
 Qy 571 GlnSerileMetTyrAspAlaThrLysAsnGlu----- 581  
 Db 2568 CAATCTATTATGTACGACGCGAAGAAATGAAATATGCGTAAAGTTTCAGAAAAATTGAA 2509  
 Qy 582 -----TyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyileGly 595  
 Db 2508 GTTTTAAATATCATATTACAG-TTCTACAAAAATTTGACACTAAATACCGAATCGGT 2450  
 Qy 596 ArgPheGluAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGlu 615  
 Db 2449 AGATTTGAAATAGCGCAACAGAGCGAAGAAATATGTTTGAACGCTCTCCCGATAAGAA 2390  
 Qy 616 GlnLysValLeuMetPheilellelleSerLysArgGlnLeuAsnAlaTyrGlyPheVal 635  
 Db 2389 CAATAAGTCTTAATGTTCATTATCATTTCCAAACGACAACTGAATGCTTACCGTTTGTG 2330  
 Qy 636 LysHisTyrCysAspHisThrileGlyValAlaAsnGlnHisileThrSerGluThrVal 655  
 Db 2329 AAACATTATGGGATCACACCATCGGTGTAGCTAATCAGCATATTACTTCTGAAACAGTC 2270  
 Qy 656 ThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgilePheTyrGlnile 675  
 Db 2269 ACAAAAGCTTGGCATCACTAAGGCACGAGAAAGGATCAAAACGAAATTTTCTATCAAT 2210  
 Qy 676 AlaLeuLysileAsnAlaLysLeuGlyileAsnGlnGluLeuAspTyrSerGluile 695  
 Db 2209 GCATTTGAAATCAACCGAAATTAGGAGGTATTAACCGAGGCTTGACTGGTCAGAAAT 2150  
 Qy 696 AlaGluileSerProGluLysGluArgArgLysThrMetProLeuThrMetTyrVal 715  
 Db 2149 GCAGAAATATCACCAGAGAAAGAAAGACGGAACCAATGCCATTAACTATGATGTT 2090  
 Qy 716 GlyileAspValThrHisProThrSerTyrSerGlyileAspTyrSerileAlaVal 735  
 Db 2089 GGAATTTGATGTAATCATCCAACTCTACAGTGGAAATGGAATTTATTTATAGCGGTGTA 2030  
 Qy 736 ValAlaSerileAsnProGlyGlyThrileTyrArgAsnMetileValThrGlnGluGlu 755  
 Db 2029 GTAGCGAGTATCAATCCAGGTGGAATATCTATCGAAATATGATGTGACTCAAGAAGAA 1970  
 Qy 756 CysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspilleLeuGluAla 775  
 Db 1969 TGTGTCCTCGGTGAGGTGAGTGGCTCATGACGGAAGAACAGATATTTTGGAGCA 1910  
 Qy 776 LysPheValLysLeuLeuArgGluPheAlaGlu----- 786  
 Db 1909 AAGTTCGTGAAATTTGCTCAGAGAAATTCGAGAAATTCGAGAAATGATGTTGATTTAAAGAT 1850  
 Qy 787 -----AsnAsnAspAsnArgAlaProAlaH 795  
 Db 1849 CTCTGGGATTTTAAATTTTTTGTAAACTTTTCAGAAACAGCAATTCGAGACCGACGCA 1790  
 Qy 795 sileValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspG 815  
 Db 1789 TATTGTAGTCTATCGAGACGGAGTTAGCGATTCGGAGATGCTACGTGTAGTCATGATCA 1730  
 Qy 815 uLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPr 835  
 Db 1729 GCTTCGATCTTTAAAAAGCGAAGTAAACAATTTCTATGTCGGAACGGGATGGAGAAATCC 1670

Qy 835 oGluProLysTyrThrPheileValleGlnLysArgHisAsnThrArgLeuLeuArgAr 855  
 Db 1669 AGAGCCGAGATACAGCTTCATTGTGATTTCAGAAAGACACAATACAGATTGCTTCAAG 1610  
 Qy 855 gMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAl 875  
 Db 1609 AATGGAAAAAGATAAGCCAGTGGTCAATAAAGATCTTACTCTGCTGAAACAGATGTCG 1550  
 Qy 875 aValAlaValLysGlnTyrPgluGluAspMetLysGluSerLysGluThrGlyileVa 895  
 Db 1549 TGTCTGCTGTTTAAACCAATGGGAGGAGGATATGAAGAAAGCAAGAAACCTGGAATGT 1490  
 Qy 895 IaenProSerSerGlyThrThrValAspLysLeuileValSerLysThrLysPheAspPh 915  
 Db 1489 GAACCCATCATCCGAAACAATGTGTGATAAATTCGTTTCGAAATACAAATTCGATTT 1430  
 Qy 915 ePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935  
 Db 1429 TTTCTTGGCATCTCATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1370  
 Qy 935 tTyrAspAspLysGlyMetSerGlnAspGluValTyr----- 947  
 Db 1369 GTATGACGATAAAGGAATGAGCCAAGATGAAGTCTATGTAAAGCGTTTGAATAGCAGTTA 1310  
 Qy 948 -----LysMetT 950  
 Db 1309 GCGATTTAGGATTTTGTAAATCCGCATATAGTTATATATAAAAAATGTTTCAGAAATGA 1250  
 Qy 950 hrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysPheProileSerLeuProValPro 970  
 Db 1249 CCTACGGAGTCTGTTTCTCTCTGCTAGATGTGAAACCCCATCTCGTTGCTGTTCCGG 1190  
 Qy 970 alHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluH 990  
 Db 1189 TTCATTATGCTCATTTATCATGTGAAAAAGCGAAAGAGCTTATCGAACTTCAAGAAC 1130  
 Qy 990 IsTyrIleGlyAspTyrAlaGlnProArgThrArgHisGluMetGluHisPheLeuGlnT 1010  
 Db 1129 ATTATCATCGTGACTATGCACAGCCAGGACTCGACAGAAATGGAACATTTTCTCCAA 1070  
 Qy 1010 hrAsnValLysTyrProGlyMetSerPheAla 1020  
 Db 1069 CTAACGTGAAGTACCTCGAATGTGTTCCGA 1038

RESULT 3

CBRG18K16  
 Caenorhabditis briggsae 41648 bp DNA linear INV 04-NOV-2000  
 DEFINITION complete sequence.  
 ACCESSION AC084520  
 VERSION AC084520.1 GI:11094970  
 KEYWORDS HTG.  
 SOURCE Caenorhabditis briggsae  
 ORGANISM Caenorhabditis briggsae  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 41648)  
 AUTHORS Washington University Genome Sequencing Center.  
 TITLE The C. briggsae Genome Sequencing Project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 41648)  
 AUTHORS Waterston, R.  
 JOURNAL Direct Submission  
 COMMENT Submitted (04-NOV-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63110, USA  
 e-mail: japieth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.



It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

## FEATURES

source  
1. .41648  
Location/Qualifiers  
/organism="Caenorhabditis briggsae"  
/mol\_type="genomic DNA"  
/strain="Gujarat G16"  
/db\_xref="taxon:6238"  
/clone="G18K16"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 41648  
Score: 2867.00 Matches: 638  
Percent Similarity: 50.7% Conservative: 147  
Best Local Similarity: 41.2% Mismatches: 203  
Query Match: 53.6% Indels: 562  
DB: 13 Gaps: 20

US-10-645-746-3 (1-1020) x CBRG18K16 (1-41648)

```
QY      6  ProGluLeuGluLys-----GlyPheTyrArgHisSerLeuAspProGlu----- 20
Db      7590 CCAAAATACAGAAAGCCAAATTAATACTACTCTCATCTTGGATCCAGT-GAGTCT 7648
QY      21 -----MetLysTrpLeuAla 25
Db      7649 TTAGACGCGCAAGAATCTAGAAAAAAATCAGAAATTCAGAGCATGAAGTGGTTATCA 7708
QY      26  ArgProThrGlyLysCysAspGlyLysPheTyrGluLysLysValLeuValAsn 45
Db      7709 AGCCCAAGAAAGAAATGCGAAGGGAAGTATTATGTGCAAAAGTAAATTTGCTGGTCAAC 7768
QY      46  TrpPheLysPheSerLysLysLeuTyrAspArgGluTyrGluTyrGluValLysMet 65
Db      7769 TGGTTCGGATTACAGCAATATTACGATCGAGAGTATTATGAGTATTATTTGGATATG 7828
QY      66  ThrLysGlu-----ValLeuAsnArgLysProGlyLys-----Pro 77
Db      7829 TACAAAGAAAGAGATATATGGATAGAGAACTAAGTCGTTCAAAACAAAGCAGATGCCA 7888
QY      78  PhePro-----LysLysThrGluLeuProle-----Pro 87
Db      7889 TATCCAAATGTAATGAAAGATAAACAATGAAGTCGTAATGTGCTTTTTCACGCCCG 7948
QY      88  AspArgAlaLysLeuPheTrpGlnHisLeuArgHisGluLysLysGlnThrAspPheIle 107
Db      7949 GAACGATCTGAAGCGTCTCGAAGCACTTGAGATTCGAGCAACAAGAAATCCGTTCCAA 8008
QY      108 LeuGluAspTyrValPheAspGluLysAspThrValTyrSerValCysArgLeu---Asn 126
Db      8009 ATTGAAGACTATGTTTGTATGAGAAGGATACGGTTTACAGTATTACAAAGACCCCAAC 8068
QY      127 ThrValThrSerLysMetLeuValSerGluLysValValLysLysAspSerGluLysLys 146
Db      8069 GGAATGCTCTCGAAT-----CCTGATCTCTCAAAACCGA 8104
QY      147 AspGluLysAspLeuGluLysLysIleLeuTyrThrMetIleLeuTyrArgLysLys 166
Db      8105 GACATCAATAGTGCTTAA-----ATGCATTACAATGTATAA 8143
QY      167 PheHisLeuAsnPheSerArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyr 186
Db      8144 TTCATCTATACCTTTAGTAGAAGATCCGACGAGGATGAAGCAAGCAAGAAAGTTGT 8203
QY      187 LysPheLeuLys----- 190
Db      8204 AAATTTTGAAGGTTCTGTATTTACTGCTAAACAAACAATATTATTCGAAAGATTTC 8263
QY      191 -AsnValMetThrGlnLysValArgTyrAlaProPheVal----- 203
Db      8264 GAATGTATTACACAAAGATTCGTTGTGACCGCTATGGCTAATGAAATCGGATTGTG 8323
```

```
QY 209 ----- 209
Db 9404 GCGAAACACCGCTTACTAGTTTGGAAAAAAATCTTTTTTTTTCGTTTTTCTAGGA 9463
QY 210 -----GlnPheAlaLysAsnPheValT 217
Db 9464 ATTTCATTTCCCGATTCCGAACAGCCGTTTTTTTCAGCCAACTTGCCAAAATTTTCATCT 9523
QY 217 YrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluG 237
Db 9524 AGCATGGAAACTCCATTTCTCCGAGTTCCAGAAATCTTTTTCAGCACCCCTGATCGATTTGAAC 9583
QY 237 InSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleValLeuG 257
Db 9584 ATTCTCTGGAATCGCCGACCGATTTGAATCTTGTTCCGAATCTATATCGCTGTTAAAG 9643
QY 257 luLeuPheAspGlyGluProValLeuAsnPheAla ----- 268
Db 9644 AACTTTTCAGCGAGACCGGTTCTCAATTTTGCAGTGAGAATTAATAATTTTAAGGCT 9703
QY 269 -----IleValAspLysLeuPheTyzAsnAlaProLysMet 280
Db 9704 GAAATATATAAACATTAATTTTCAGTTATTGACAAATTTCTTACAAAGCTCCACAAATG 9763
QY 281 SerLeuLeuAspTyzLeuLeuLeuIleValAspProGlnSerCysAsnAspValArg 300
Db 9764 TCTCTCTCGACTACTCTGTTCTGATTTGGATTCGCAATCTCATATGGACGAAGTTGCA 9823
QY 301 LysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgPro 320
Db 9824 AGAAACCATAAAGAACAGTTGAATTCGGGAATTTTGACAAATTAAGCCAGCTATGTTTCAA 9883
QY 321 ArgIleArgGlnLeuLeuLeuLeuLysLeuLysCysAlaGluValTrpAsnGlu 340
Db 9884 AGCATTGATCAATTGCTTCGGAATTTAAAGTTAAAGAGCAGCATCGTATGGAATCCTAAA 9943
QY 341 -----MetSerArgLeu 344
Db 9944 AGGTATGAATCTACTATTGGAAGATGAGAGTTAATCAAAATTTATTCAGGTACAGTTT 10003
QY 345 ThrGluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyzLysVal 364
Db 10004 GCTGAACGTCATCTGACATTTATCAGATTGAGCAAGTATAATTTCCACGAAGAAATGATT 10063
QY 365 ThrGlyLysSerAspArgGlyArgAsnAlaLysLysTyzAspThrThrLeuPheLysIle 384
Db 10064 CCAGTTCCAAGAGGAGACAGACAGAAATGCCCAATTCAGAGTTCCCATTTGTTCAAAT 10123
QY 385 TyrGluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSer 404
Db 10124 TATGAGAGAACCGTAAATACATCGAGTTTCTCGCTTACCATTGGTGTGTGAAGAC 10183
QY 405 GlyAlaLysGluTyzAlaValProMetGluHisLeuGluValHisGluLysProGlnArg 424
Db 10184 CGTAAATACAGATATTTCGGTCCCATGGAGTTTCTAGAAAGTGATACAAAAACCAACGC 10243
QY 425 TyrLysAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLys 444
Db 10244 TACAAAATTCGAATCCAGCTTTGCGATGCCAAGACAAATTTGTGAATGCAGCAGACGCGAT 10303
QY 445 ProHisAspTyzLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGlu 464
Db 10304 CCTCATATGTATAGAGGAGGAGCGCTTGAGATGTTGAAGATCTCGATTTTTCATCTGA 10363
QY 465 GluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysPro 484
Db 10364 TCGCTGAATCTTGTTCAACGGTTTGGATTCAGTACCGATCTGAAAATGATAGAGTGTGT 10423
QY 485 GlyLysValIleLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThr 504
Db 10424 GGAAGAGTCTGAAGAAACCATCCTCGTCAACAAAGATAATCAAAAGATCAGTATGACA 10483
QY 505 ProValIleArgGlyPheGlnGluLysGlnLeuAsnValValProGluLysGluLeuCys 524
```

---

```
10484 CTTGGTTAGGATTCCAGAGAAAGCGTTGAACGTGGTTCCAGAAAAGAACTCTGT 10543
QY 525 CysAlaValPheValValAsnGlu---ThrAlaGlyAsnProCysLeuGluGluAsnApp 543
10544 TGTGCCATTTTGTCTTCGCGAGCAAAAGACAGAGCCATGTTTGAAGAAGAAGAT 10603
QY 544 ValValLysPheTyzThrGluLeuIleGlyCysLysPheArgGlyIleArgIleGly 563
10604 GTGAGGAGTCTTACAAGACGCTCATCGATGTTGTAATTCAGATCTATTGAAATCGGA 10663
QY 564 AlaAsnGluAsnArgGlyAlaGlnSerIleMetTyzAspAlaThrLysAsnGluTyzAla 583
10664 AAACATCAAAATAGTCAAGTTCTGTTCTGTTGCAAGATCCTGAAAGTGGGAAGTATGG 10723
QY 584 -----PheTyzLysAsn 587
10724 GTGAGTTGGTTTATCTCGAGTTAGTTACTGATATCCATTTATTTTTCAGTTCTACCCGAAC 10783
QY 588 CysThrLeuAsnThrGlyIleGlyArgPheGluIleAlaAlaThrGluAlaLysAsnMet 607
10784 GTGAGCTCATCTGCGGGCGTACGCAACTTTTAGAGCTGTGCTAACGATCGGAAGCAATG 10843
QY 608 PheGluArgLeuProAspLysGluGlnLysValLeuMetPheIleIleLysSerLysArg 627
10844 TTTGAGAGACTTGTCTGACAGAATAACAAGATTCTACTGTTTCTGTTTTCGAAACGT 10903
QY 628 GlnLeuAsnAlaTyzGlyPheValLysHisTyzCysAspHisThrIleGlyValAlaAsn 647
10904 CGTTGGAATTTCTATGTTTCTGTTGAAACAGTTCTGCGATGTTGAACTGTTGTTAGCAGT 10963
QY 648 GlnHisIleThrSerGluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGly 667
10964 CAACATGTTTACTGCCACGTCGCGAAAGAGCTCTACAGGATATGTTCTCTCATAACTT 11023
QY 668 SerLysArgIlePheTyzGlnIleAlaLeuLysIleAsnAlaLysLeuGlyIleAsn 687
11024 TCTAAACGAATTTCTATCAATCGCTTTAAAGATCAATGGAATACTAGGAGGTGTAAT 11083
QY 688 GlnGluLeuAspTyzSerGluIleAlaGluIleSerProGluGluLysGluArgLys 707
11084 CAAGAATCGATTTGTTGCGAAATTCAGAGATGACATGTTGAAAGAAAAGAAAGGAAG 11143
QY 708 ThrMetProLeuThrMetTyzValGlyIleAspValThrHisProThrSerTyzSerGly 727
11144 AATATGCCACTCCGAATGTATGTTGGATTTGATGTCATCATCCGACGAATGGAATGGA 11203
QY 728 IleAspTyzSerIleAlaAlaValValAlaSerIleAsnProGlyGlyThrIleTyzArg 747
11204 ATCGACTACTCGATAGTCTATTTGCGCAAGTATGATGATCCAGGAGAAACCGTTATCGT 11263
QY 748 AsnMetIleValThrGlnGluLysCysArgProGlyGluArgAlaValAlaHisGlyArg 767
11264 AATATGATTTGTAACCTCAAGAGAGAAATCGTCCAGGTGAAACGTCGCTAGTCTCATGGAAG 11323
QY 768 GluArgThrAspIleLeuGluAlaLysPheValLysLeuLeuArgGluPheAlaGlu--- 786
11324 GAGAGAACTGATATCTTGGACGGAAATTCGTGAAAGCTTTTCGAGATCTTTGACAGAGTA 11383
QY 787 -----AsnAs 788
11384 TTAATCTTCTTATCAACAGCGCAACATAGTTTGTGTTCTTATTTGTTTACAGAACAA 11443
QY 788 AspAsnArgAlaProAlaHisIleValValTyzArgAspGlyValSerAspSerGluMe 808
11444 CGAAAACCGTATTTCCCTCATATGTTGTTTATCGAGACGGTGTGAGGACCTCAGAAAT 11503
QY 808 tLeuArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSe 828
11504 GCTGGTGTCTGATCATGATGAATTAAGGTCAATGAAAGCGGAAGTTGAACGATTCCTGAA 11563
QY 828 rGluArgAspGlyGluAspProGluProLysTyzThrPheIleValIleGlnLysArgHis 848
```



Pred. No.: 8 53e-57 Length: 4148  
 Score: 582.00 Matches: 238  
 Percent Similarity: 40.2% Conservative: 172  
 Best Local Similarity: 23.3% Mismatches: 370  
 Query Match: 10.9% Indels: 243  
 DB: 5 Gaps: 38

US-10-645-746-3 (1-1020) x BC063275 (1-4148)

QY	21	MetLysTrpLeuAlaArgProThrGlyLysCysAspGlyLysPheTyroGlu-----	37
DB	203	ATGGGATGAA-GGGGACCTCGGAGCAGCTCGGGCGCTTACCTGCCCCCTGCAG	261
QY	38	-----LysLysValLeuLeu 42	
DB	262	CAGGTGTTCCAGCACCTCCGGGCTGGCATTTGGCAGCTGTGGGAAACCAATCAAGCTC	321
QY	43	LeuValAsnTrpPheLysPheSerLysLysIleTyroAspGluTyroGluTyroGlu 62	
DB	322	CTGGCCNATTTCTTGAGGTGGACATCCCTAAGATCGAGCTGACCATCAGGGTGGAC	381
QY	63	ValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysThr 82	
DB	382	ATCAAGCCGATAAG-----TGTCCTCGTAGAGTC 411	
QY	83	GluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeuArgHisGluLysLys 102	
DB	412	AACCGGAAGTGGTGGAA-----TACATGGTCCAGCATTTCAAGCCT 453	
QY	103	GlnThrAspPheIleLeuGluAspTyr-----ValPheAspGluLysAspThrValTyr 120	
DB	454	CAG-----ATCTTGGTGTATCCAGACCTGTGTATGATGGAAGAAGAACATTATAC	504
QY	121	SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysVal-----	138
DB	505	-----ACTGTACAGCACTGCCCATTTGGCAACGAAAGCGGTGCACTTT 546	
QY	139	---ValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysLysLys 157	
DB	547	GAGGTGACAAATCCCTGGGGAAGGAGGAT---CGAATCTTTAAGTCTCCATCAAGTGG	603
QY	158	ThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGlu 177	
DB	604	CTAGCATTTGTGAGCTGGGAATGCTGCATGAGGCCCTGTGTCAGCGCCAGATCCCTGT	663
QY	178	LysAspGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysVal 197	
DB	664	CCCTTGAGTCTGTCAAGCCCTGATGTGGCCATGAGGCACCTGGCATCC-----ATG	717
QY	198	ArgTyrAlaPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyr 217	
DB	718	AGGTACACCCCTGTG-----	732
QY	218	AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln 237	
DB	733	---GGCGCTCTCTCTCACCCTGAGGCTACTACACCCG-----	774
QY	238	SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyLysGlu 257	
DB	775	-----CTGGGGGTGGGCGGAGGTCTGTTCTGGCTTTCACCCAGTCTGTGGCCCT	825
QY	258	LeuPheAspGlyLeuProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAla 277	
DB	826	GCCATG---TGAAGATGATGTCTCAACATTTGATGTCTCAGCCACTCCCTTTTATAGGCA	882
QY	278	ProLysMetSerLeuLeuAspTyrLeuLeuIleValAspProGlnSerCysAsnAsp 297	
DB	883	-----CAGCCAGTGTGATGTTGATGTGAGGTCTGACATCAGAACATTAATGAG	936
QY	298	AspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAla 317	
DB	937	CAGCCCAAG-----CCCCCTCAGGACTCT 960	

QY	318	AlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLys-----CysAla 334	
DB	961	CAGCGCGTTCGCTTCCACAGAGATCAAGGCCCTGAGGTGGAAGTCAACCCACTGTGGA	1020
QY	335	GluValTrpAspAsn-----GluMetSerArgLeuThrGluArgHisLeuThr 350	
DB	1021	CAGATGAAGAGGAAGTACCGCGTGTGTAATGTATTACCGCTCGCCCTAGCCATCAGACA	1080
QY	351	PheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLysSerAspArg 370	
DB	1081	TTCCCTTTACAGCTGAGAGTGGACAGACTGTG-----	1113
QY	371	GlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGluGluAsnLysLys 390	
DB	1114	-----GAGTGCACAGTGGCACAGTATTTCAGCAGAAATATAAC 1152	
QY	391	Phe-----IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTyr 409	
DB	1153	CTTCAGCTCAAGTATCCCATCTGCTCCCTTACAAGTTGGCCAGGAACAAAAGCATACC	1212
QY	410	AlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArgIle 429	
DB	1213	TACCTTCCCTAGAGTCTGTAAACATT---GTGGCTGGCAGCGCTGTATTAAAAAGCTG	1269
QY	430	AspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLys 449	
DB	1270	ACCGACACACAGACCTCGACCATGATAAGGCCACAGCTAGATCCGCTCCAGACAGACAG	1329
QY	450	GluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheVal 469	
DB	1330	GAGGAGATCAGTCGCTGATGAAGAAATGCCAGCTACAATCTAGATCCC-----TACATC	1383
QY	470	GluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys 489	
DB	1384	CAGGAATTTGGATCAAAAGTGAAGATGACATGACGAGGTGACAGGGGAGTGTGCTGCCG	1443
QY	490	GluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro-----	505
DB	1444	GGCCCATCTTGCAGTACGGCGCGGAACCGGGCCATTTGCCACACCCCAATCAGSGTGTC	1503
QY	506	---ValIleArgGlyPheGlnGluLysGlnLeu---AsnValValProGluLysGluLeu 523	
DB	1504	TGGGACATGCGGGG-----AAACAGATTCTACAATGGGATTCAGATCAAAAGTCTGG	1554
QY	524	CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluAsnAsp 543	
DB	1555	GCATCGCTGCTTCGACCCCAAAA-----CAGTGTGCGAAGAG-----	1596
QY	544	ValValLysPheTyrThrGluLeuIle-Gly-----GlyCysLysPhe 557	
DB	1597	GTGCTCAAGAACTTCACAGACCCAGCTGCGGAAGATTTCCAAAGATGCGGGGATGCCTATC	1656
QY	557	eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAl 577	
DB	1657	CAGGCTCAACCTTGTCTTCTGCAAAATATGCACAGGGGGCA-GACAGCGTG-----	1704
QY	577	aThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPh 597	
DB	1704	-----	1704
QY	597	eGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLys 617	
DB	1705	-----GAGCTATGTTCCGGCATCTCTCAAGAACACCTACTACTCAGG 1742	
QY	617	sValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHis 637	
DB	1743	GCTGAGCTCATTTATGTCATCTCTCCAGGGAAGACCGCGGTGTATGCTGAGGTGAACG	1802
QY	637	sTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLys 657	
DB	1803	TGTCGAGATACACTCTTGGGAATGGCTACGAGGTGTGTGTCAGGTGAAGAACGTGGTCAA	1862
QY	657	sAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLe 677	



QY	198	ArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyr	217	QY	544	ValValLysPheTyrThrGluLeuIle-Gly-----GlyCysLysPhe	557
DB	724	AGGTACACCCCTGTG-----	738	DB	1603	GTGCTCAAGAACTTCACAGACAGCTGCGGAAGATTTCAGAGATGCGGGATGCTCATC	1662
QY	218	AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln	237	QY	557	eArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIleMetTyrAspAl	577
DB	739	-----GGCCGCTCTCTTCTTCACCGCTGAGGGCTACTACCAACCCG-----	780	DB	1663	CAGGTCACACTTGTCTGCAATATGCACAGGGGCA-GACAGCGTG-----	1710
QY	238	SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyIleLysGlu	257	QY	577	aThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGlyArgPhe	597
DB	781	-----CTGGGGGTGGCGCGAGGTCTGTTCCGGCTTTCACCAAGTCTGTGCGCCCT	831	DB	1710	-----	1710
QY	258	LeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAla	277	QY	597	eGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGluGlnLys	617
DB	832	GCCATG---TGGAAGATGATGCTCAACATTGATGTCTCAGCCACTGCCCTTTTATAGGCA	888	DB	1711	-----GAGCTATGTTCCGGCATCTCAAGAACACCTACTACTCAGG	1748
QY	278	ProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGlnSerCysAsnAsp	297	QY	617	eValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheValLysHl	637
DB	889	-----CAGCCAGTGATTGAGTTCAATGTGTGAGGTCTGGACATCAGGAACATAGATGAG	942	DB	1749	GCTGCGACTCATTTATGTCATCTCCAGGGAAGACGCGGTGTATGCTGAGTGAACG	1808
QY	298	AspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThrIleArgGlnAla	317	QY	637	eTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrValThrLys	657
DB	943	CAGCCCAAG-----CCCTCAGCGACTCT	966	DB	1809	TGTCGAGATACACTCTTGGGAATGCTACGCACTGTGTGAGGTGAAGAACGTGTGTCAA	1868
QY	318	AlaArgProArgIleArgGlnLeuLeuAsnLeuLysLeuLys-----CysAla	334	QY	657	eAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIleAlaLe	677
DB	967	CAGCGGTTTCGTTTCCACCAAGAGATCAAGGGCTGAAGGTGGAAAGTCAACCCACTGTGGA	1026	DB	1869	GACC-----TCACCTCAGACTCTGTCCAACTCTGCGCT	1901
QY	335	GluValTrpAspAsn-----GluMetSerArgLeuThrGluArgHisLeuThr	350	QY	677	uLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTrpSerGluIleAlaGl	697
DB	1027	CAGTGAAGAGGAAGTACCGCTGTGTAATGTTTACCCGTGCGCTGCTAGCCATCAGACA	1086	DB	1902	CAAGATCAATGTCAAACTTGGTGCAATTAACAACATCCTA-----	1941
QY	351	PheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGlyLysSerAspArg	370	QY	697	uLysSerProGluGluLysGluArgLysThrMetProLeuThrMetTyrValGlyIle	717
DB	1087	TTCCCTTACGTGGAGGTGGACAGACTGTG-----	1119	DB	1942	-----GTCCACACACCGCTCTGCCGCTTTTCAACAGCCAGTG---ATATTCTGGGAGC	1994
QY	371	GlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGluGluAsnLysLys	390	QY	717	eAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAlaAlaValAla	737
DB	1120	-----GAGTGCAAGTGGCACAGTATTTTCAAGCAGAAATATAAC	1158	DB	1995	AGATTTACACACCCCGCAGCGGGGTGGGAAAAAACCCTTATATCAGCAGTGGTAGG	2054
QY	391	Phe---IleGluPheProHisLeuProLeuValLysValLysSerGlyAlaLysGluTyr	409	QY	737	aSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGlnGluCysAr	757
DB	1159	CTTCAGCTCAGATATCCCATCTGCCCTCCTACAGTTGGCCAGGAACAAAGATACC	1218	DB	2055	CAGTATGATGCCACCCCGCAGCGCATCTGTCTACTGTGCGGTACAG-----	2103
QY	410	AlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLysAsnArgIle	429	QY	757	gProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGlu-----Al	775
DB	1219	TACCTTCCCTAGAGTCTGTAACTT---GTGGCTGGCGAGCGCTGTATTAAAAAGCTG	1275	DB	2104	-----CGACCACCGCAAGAGATCATTTGAAGACTTGTTC	2135
QY	430	AspLeuValMetGlnAspLysPheLeuLysArgAlaThrArgLysProHisAspTyrLys	449	QY	775	aLysPheValLys---LeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProAl	794
DB	1276	ACCGCAACACACCTCGACCATGATAAAGGCCACAGCTAGATTCGCTCCAGACAGACAG	1335	DB	2136	CTACATGTGCGTAGCTCTCTCATCCAATTTACAGTCCACCCCGTTTCAAG---CCTAC	2192
QY	450	GluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGluLeuAsnPheVal	469	QY	794	aHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAs	814
DB	1336	GAGGAGATCAGTCGCTGTAGAGATGCGACACTACACTTAGATCCC-----TACATC	1389	DB	2193	CGCATCATCTTCTACCGAGATGGGTGCTCAAGGCCAGCTACCCAGATATCTCCACTA	2252
QY	470	GluArgPheGlyLysSerLysLeuGlnMetIleGluCysProGlyLysValLeuLys	489	QY	814	pGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAs	834
DB	1390	CAGGAATTTGGGATCAAAAGTGAAGATGACATGACGAGGTGACAGGGCGAGTGTGCGG	1449	DB	2253	TCAGCTACTGCCCATTCGTGATGCCCTGCATCAAACTG-----GAAAGGA	2297
QY	490	GluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro-----	505	QY	834	pProGluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuAr	854
DB	1450	GCGCCCATCTTCAGTACGCGCGCGGAACCGGCCATTTGCCACACCCCAATCAGGGTGT	1509	DB	2298	CTACCGCTGGGATCATTTATTTGTGTGCGAGAAACGCCATCACACCGCTTTC--	2355
QY	506	---ValIleArgGlyPheGlnGluLysGlnLeu---AsnValValProGluLysGluLeu	523	QY	854	gArgMetGluLysAspLysProValAlaAsnLysAspLeuThrProAlaGluThrAspVa	874
DB	1510	TGGGACATCGGGGG-----AAACAGTTCTACAAATGGGATTTGAGATCAAAAGTCTGG	1560	DB	2356	-TGCTGCAAGAATGACGGAATTTGGGAAGAGTGTAACTCCAGCT-----	2403
QY	524	CysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluAsnAsp	543	QY	874	lAlaValAlaAlaValLysGlnTrpGluLysAspMetLysGluSerLysGluThrGlyIle	894
DB	1561	GCCATCGCTGCTTCCACCCCAAAAA-----CAGTGTGAGAGAGAG-----	1602	DB	2403	-----	2403
				QY	894	eValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAs	914



Db 2404 -----GGGACCAAGTGGACCAACATACCCACCCTTGGATTGA 2447  
 Qy 914 pPhePheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGlyHisThrVa 934  
 Db 2448 CTTCTATCTGTGCAGCCAGCAGCATCCAGGCGACCAAGCCATCCCTACTATGT 2507  
 Qy 934 lMetTyrAspAspGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAl 954  
 Db 2508 TCITGGGATGACAAACCGTTTACAGCAGATGAGCTCCAGATCCGAGTACCAAGCTGTG 2567  
 Qy 954 aPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHi 974  
 Db 2568 CCACACTTACGTAGTGCACACGCTCTGTCTATCCAGCACCTGCTCTACTATGCGG 2627  
 Qy 974 sLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHisTyrIleGlyAs 994  
 Db 2628 CCTGGTGGCTTTCGGGACGGA---TACCACCTGGTGACAGAGCATGACAGTGGAGA 2684  
 Qy 994 p 994  
 Db 2685 G 2685

## RESULT 6

AF093097 7478 bp mRNA linear PRI 16-NOV-1999  
 LOCUS Homo sapiens putative RNA-binding protein Q99 mRNA, complete cds.  
 AF093097  
 DEFINITION AF093097.1 GI:6002622

## SOURCE

Homo sapiens (human)  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

## REFERENCE

1 (bases 1 to 7478)  
 Koesters, R., Adams, V., Betts, D., Moos, R., Schmid, M., Siermann, A.,  
 Doeberitz, M., and Briner, J.

Hassem, S., Weitz, S., Lichter, P., Heitz, P.U., von Knebel

Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,  
 genomic organization, localization to chromosomal bands 1p34-p35,  
 and expression

Genomics 61 (2), 210-218 (1999)

10534406

2 (bases 1 to 7478)

Koesters, R., Briner, J., Moos, R., Schmid, M., Doeberitz, M.V.K. and  
 Betts, D.

Direct Submission

Submitted (18-SEP-1998) Surgery, University Hospital, INF 110,  
 Heidelberg 69120, Germany

Location/Qualifiers

1. 7478  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosomes="1"  
 /map="1p34-p35"  
 214..2787  
 /notes="similar to Arabidopsis thaliana AGO1"  
 /codon\_start=1  
 /product="putative RNA-binding protein Q99"  
 /protein\_id="AA00068.1"  
 /db\_xref="GI:6002623"

## CDS

/translation="MSAGPSGAAGAYLPPLQQVFOAPRRPGITVGKPKILANYFE  
 VDIPKIDVHYEDIKPDKPRVRNREVYVQHPKQIFGDRKVPYDGKNIYTVT  
 ALPIGNREVDFVTIPEGKDRIFKVISIKWLAIVSRMLHALVSGQIPVLESVOAL  
 DYAMRHLSMRVTPVGRSFPSPGEGYHPLGGREWFPGHQSVPRAKMWMLNDVS  
 ATAFYKAQVIFEMCEVLDIRNIDEQPKLDSQRVRFTGKINGLVKVEYTHCGQMRK  
 YRVNVRPASHQTPPLQESGQVECTVAQYFKOKYNLQKLYPLHPLCLQVGOGRH  
 TYPLEVCNVASQRCIKKLTQNTSTMIKATARSAPDQEEISRLMKQASNYLDPYI  
 QBFIKQKDDTEVTGRVLPAPLQYGGNRRAITPNQGWDMRGKQFYNGEIKWA  
 IACFAPQKCRREVLNFTDLRKISKDAGMPTQGPCFCYKAGQADSVPMFHLKN  
 TYSGLQLIIVLPKTPVYAEVKRVGDTLLGNATQCVQVNVNVTSPQTLNLCILKN

## ORIGIN

Alignment Scores: 7.87e-56 Length: 7478  
 Pred. No.: 577.00 Matches: 237  
 Score: 577.00 Conservative: 173  
 Percent Similarity: 40.2%  
 Best Local Similarity: 23.2%  
 Query Match: 10.8%  
 DB: 5 Gaps: 38

US-10-645-746-3 (1-1020) x AF093097 (1-7478)

Qy 21 MetLysTyrLeuAlaArgProThrGlyLysCysAspGlyLysPheTyrGlu----- 37  
 Db 209 ATGGGATGGAA-CGGGACCCCTCGGAGCAGCTGCGGCGCTTACCTGCCCCCCTGCAG 267  
 Qy 38 -----LysLysValLeuLeu 42  
 Db 268 CAGGTGTTCCAGCACCTCGCCGCTGGCATTTGGCAGCTGTGGGAAACCAATCAAGCTC 327  
 Qy 43 LeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGluTyrGluTyrGlu 62  
 Db 328 CTGGCCAATTACTTTGAGGTGGACATCCCTAAGATCGACGTGTACCACCTACGAGGTGAC 387  
 Qy 63 ValLysMetThrLysGluValLeuAsnArgLysProGlyLysProPheProLysLysThr 82  
 Db 388 ATCAAGCCGGATAG-----TGTCCCGCTAGATGC 417  
 Qy 83 GluIleProIleProAspArgAlaLysLeuPheTyrGlnHisLeuArgHisGluLysLys 102  
 Db 418 AACCGGAGAGTGTGGAA-----TACATGTCACGACATTTCAAGCTT 459  
 Qy 103 GlnThrAspPheIleLeuGluAspTyr-----ValPheAspGluLysAspThrValTyr 120  
 Db 460 CAG-----ATCTTTGTTGATCGCAAGCCCTGTGTATGATGMAAGAAAGAACATTAC 510  
 Qy 121 SerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLysVal----- 138  
 Db 511 -----ACTGTACAGCAGCTGCCCATTTGGCAACGACGCGTGCAGCTTT 552  
 Qy 139 ---ValLysLysAspSerGluLysLysAspGluLysAspLeuLysLysIleLeuTyr 157  
 Db 553 GAGGTGACAAATCCCTGGGGAAGGAAGGAT---CGAATCTTTAAGGTCTCTCAACAGTGG 609  
 Qy 158 ThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsnProGlu 177  
 Db 610 CTAGCCATTGTGAGCTGCGCAATGTCATGAGGCGCTGTGTGTCAGCGCCAGATCCCTGTT 669  
 Qy 178 LysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGlnLysVal 197  
 Db 670 CCTTGGAGTCTGTGCAAGCCCTGGATGTGGCATGAGGACACCTGCGCATCC-----ANG 723  
 Qy 198 ArgTyrAlaProPheValAsnGluGluIleLysValGlnPheAlaLysAsnPheValTyr 217  
 Db 724 AGGTACACCCCTGTG----- 738  
 Qy 218 AspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsnArgPheGluGln 237  
 Db 739 ---GGCGCTCTCTTCTTCTCAGCCCTGAGGGCTACTACACCCG----- 780  
 Qy 238 SerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIleGlyIleLysGlu 257  
 Db 781 -----CTGGGGGTGGGCGCGAAGTCTGGTTCGGCTTTCACCAAGTCTGTGCGCCCT 831  
 Qy 258 LeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeuPheTyrAsnAla 277  
 Db 832 GCCATG---TGGAAAGATGATGCTCAACATTGATGTCTCAGCCACTGCTCTTTTATAAGGCA 888

VKLGGINNLLVPHQSAVFOQVIFLGADVTHPPAGDGKKPSITAVVGSMDAHPSRVC  
 ATVRQRPQREIIIESYVRELLIQFYKSTRFKPTRIIIFYRDGVPGQIPQILHNYH  
 LAIRDACIKLEKDYQFQITVIVVQRRHTRLFCKNKRINIGKSNIPAGTVDNTH  
 PFEPFVLCSHAGIQGTSRPSHYVLMDDNRFTADELQILYQLCHTYVTRCTSVSIP  
 APAYVARLVAFRARIYHLVDKEHDSGEGSHISGQSGNRDPQALAKAVQVHQDTRLTMYF  
 A"





Db	2628	CCTGGTGGCTTTCGGGACAGA---TACCACCTGGTGGACAAGGAGCATCACAGTGGAGA	2684
Qy	994	G 994	
Db	2685	G 2685	
RESULT 7			
LOCUS	AR439853	2910 bp	linear
DEFINITION	Sequence 105 from patent US 6664446.		
ACCESSION	AR439853		
VERSION	AR439853.1	GI:42665818	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 2910)		
AUTHORS	Heard, J., Broun, P., Riechmann, J. L., Keddie, J., Pineda, O., Adam, L., Samana, R., Zhang, J., Yu, G.-L., Ratcliffe, O., Pilgrim, M., Jiang, C.-Z., and Reuber, L.		
TITLE	Transgenic plants comprising polynucleotides encoding transcription factors that confer disease tolerance		
JOURNAL	Patent: US 6664446-A 105 16-DEC-2003;		
FEATURES	Mendel Biotechnology, Inc.; Hayward, CA		
source	location/Qualifiers		
	1. 2910		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Alignment Scores:			
Pred. No.:	2.95e-56	Length:	2910
Score:	575.50	Matches:	237
Percent Similarity:	39.0%	Conservative:	159
Best Local Similarity:	23.3%	Mismatches:	337
Query Match:	10.8%	Indels:	283
DB:	2	Gaps:	41
US-10-645-746-3 (1-1020) x AR439853 (1-2910)			
Qy	38	LysLysValLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGlu	57
Db	472	AAGAAAGTCATGGTTCGTGGCAT-----CATTTCTGTGTCAGTTGCTGATCGTAT	525
Qy	58	TyTyGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysPro	77
Db	526	CTCTACCATTCAGATGTTTCGATCAATCTCGAGGTTATATCAAG-----	570
Qy	78	PheProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeu	97
Db	570	-----	570
Qy	98	ArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyrValPheAspGluLysAsp	117
Db	570	-----	570
Qy	118	ThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSerGluLys	137
Db	571	-----ACAGTGAACAGAAACGTGATGAACTTCTGGTTAAGAAT---	609
Qy	138	ValValLysLysAspSerGlu-----LysLysAspGluLysAspLeuGluLysLys	154
Db	610	-----TATAAAGACTCTCACTTGGGAGGAAGTCACCAAGCGTATGAT---GGAAGGNA	660
Qy	155	IleLeuTyrThrMet-----IleLeuThrTyrArgLysLysPheHisLeuAsnPheSer	172
Db	661	AGCCTCTACTCTGGTCCATTACCTTTTGACTCGAAGAGTTGTTGTTGATCTGGCG	720
Qy	173	ArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyrLysPhe---LeuLysAsn	191
Db	721	GAGAAAGAGCTGACGGTTCCTCTGGGAAGACAGACCGTTTAAAGTTGCTGTAAGAAT	780



VDIPKIDVYHVEDIKDKPCRRVNVREVMVQHPKPOIFGDRKPYVDGKNIYTVT  
ALPIGNSERVPEVTIPGEGKDIRIFKVSIKMLAIVSWRLHEALVSGQIPVPLESVOAL  
DVAMRHLSMRYTPVGRSPFPEGYHPLGGREVRVFGHQSVRPAVMKMLNIDYS  
ATFYKAQPIEFCEVBLDIRINDEQPKPLTDSQRVFTKEIKGLKVEVTHCGMKDK  
YRVNVRPASHQTFPLQLESQTVCTVAQKFKQKYNLQKYPHLPCLOVQGEORX  
TYLPLECNIVAGQRCIKLTDNOSTMIKATARSAPDROEISRLMKNASCNLDPIYI  
QERGIKVDMDTEVTGRVLPALLOYGGRNRAIATPQGVNDMRGKQFYNGIEIKWA  
IACFAPQKQREEVLNKFTDQLRKISKDAGMPIQGGPCFKYAGADSVPFMRHLKN  
TYSGLQIILVLPKTPYAEVRVKGDTLLGMATQCVQVKNNAVKTSPQTLNCLKLN  
VKLGNNIILVPHORSVAVFOOPVIFLGADVTHPAGDGKPSITAVVGSMDAHPRYC  
ATVRORPROEIIEDLSYMYRELLIOFYKSTRFKPRIIFRVDGVPGOLPOLIHLVE  
LATRDACIKLEKDYOGITVIVQKRHHHTLFCADKNSRIGKSGNI PAGTVDNTIHL  
PERFDVLCSHAGIQGSRPSHYVLDWDRNRPFADELQILITLYQLCHTYVTRCTSVSIP  
APAYARLVAFRARIYHLVDKHDHSGEGSHISGSGNDRPDQALAKAVQVQDITLRTMP  
A\*

## ORIGIN

Alignment Scores:  
Pred. No.: 8.35e-56 Length: 2574  
Score: 571.00 Matches: 232  
Percent Similarity: 40.5% Conservative: 167  
Best Local Similarity: 23.6% Mismatches: 365  
Query Match: 10.7% Indels: 222  
DB: 6 Gaps: 37

US-10-645-746-3 (1-1020) x AB081471 (1-2574)

QY 38 LysLysValLeuLeuLeuValAsnTrpPheLysPheSerSerLysLysIleTyrAspArgGlu 57  
|||  
:::  
:::  
Db 100 AAACCAATCAAGCTTCTGGCCATTACTTTGAGGTGGACATCTCTAAGATTGACGTTAC 159  
QY 58 TyrTyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysPro 77  
|||  
:::  
:::  
Db 160 CATTACGAGGTGGACATCAAGCCGGATAAG----- 189  
QY 78 PheProLysLysThrGluIleProIleProAspArgAlaLysLysLeuPheTrpGlnHisLeu 97  
|||  
:::  
:::  
Db 190 TGTCCTCGCAGATCAACCGGAGGTGGTGGA-----TACATGGTC 231  
QY 98 ArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyr-----ValPheAspGlu 115  
|||  
:::  
:::  
Db 232 CAGCATTTCAACCGCAG-----ATCTTTGGGATCGCAAGCCTGTGTATGATGGA 282  
QY 116 LysAspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSer 135  
|||  
:::  
:::  
Db 283 AAGGAGAATATTAC-----ACGGTCACAGCATGCCCATTTGGCAAT 324  
QY 136 GluLysVal-----ValLysLysAspSerGluLysLysAspGluLysAspLeuGlu 152  
|||  
:::  
:::  
Db 325 GAGAGGTTGACTTTGAGGTGACATCCCGGGAGGGGAGGAT---AGNATTTTAAG 381  
QY 153 LysLysIleLeuTyrThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSer 172  
|||  
:::  
:::  
Db 382 GTCTCCATCAAGTGGCTAGCATTTGAGCTGGCGCATGCTGCGATGAGCCTTGGTCAGT 441  
QY 173 ArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyrLysPheLysLysAsnVal 192  
|||  
:::  
:::  
Db 442 GGCAGATCCCTGTGCCCTGGAGTCTGTGCAAGCCCTGGATGGCCATGAGGCACCTG 501  
QY 193 MetThrGlnLysValArgTyrAlaProPheValAsnGluLysValGlnPheAla 212  
|||  
:::  
:::  
Db 502 GCATCT-----ATGAGGTACACCCCTGTG----- 525  
QY 213 LysAsnPheValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspPro 232  
|||  
:::  
:::  
Db 526 -----GSCCGCTCTCTTCTTCTACCGCTGAGGGCTACTACACCCG 567  
QY 233 AsnArgPheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyr 252  
|||  
:::  
:::  
Db 568 -----CTGGGGGTGGGGCGAGGTCTGGTCTGGCTTTCC 603  
QY 253 IleGlyIleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLys 272  
|||  
:::  
:::  
Db 1498 -----GAGCCCATGTTCCGGCATCTCAAGAA 1523

Db 604 CAGTCTGTGGCCCTCGCCATG---TGAAGATGATGTCAACATTGATGTCTCAGCCACT 660  
QY 273 LeuPheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuLeuValAspPro 292  
|||  
:::  
:::  
Db 661 GCCTTCTCAAGACA-----CAGCCAGTGAATGAGTTCATGTGTGAGGTCTTGACATC 714  
QY 293 GlnSerCysAsnAspAspValArgLysAspLysLysThrLysLeuMetAlaGlyLysMet 312  
|||  
:::  
:::  
Db 715 AGGAACATAGACGAACCAACCAAG----- 738  
QY 313 ThrIleArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLys 332  
|||  
:::  
:::  
Db 739 CCCCTCAGGATTTCCCAAGCGTGTTCGTTTACCAAGAGATATAAAGGCCCTCAAGGTGAA 798  
QY 333 -----CysAlaGluValTrpAspAsn-----GluMetSerArgLeuThr 345  
|||  
:::  
:::  
Db 799 GTGACCCACTGTGGACAGATGAAGAGAAATACCGCGTGTGTATGTATCCCGCGCCCT 858  
QY 346 GluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThr 365  
|||  
:::  
:::  
Db 859 GCTAGCCATCAGACGTTCCCTTTCAGCTGGAGAGTGGACAGACTGTGGAGTGCACGTG 918  
QY 366 GlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyr 385  
|||  
:::  
:::  
Db 919 GCACAGCATTTCAAG-----CAGAAATATAACCTTCAGCTC----- 954  
QY 386 GluGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGly 405  
|||  
:::  
:::  
Db 955 -----AAGTATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 993  
QY 406 AlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyr 425  
|||  
:::  
:::  
Db 994 CAGAAGCATACCTATTGTCCTCGAGTCTGTAAACATT---GTGGCTGGCAGCGGTGC 1050  
QY 426 LysAsnArgIleAspLeuValMetGlnAspLysPheLysLysArgAlaThrArgLysPro 445  
|||  
:::  
:::  
Db 1051 ATTAAGAGTACTGACCAACAGACTTCAACCATGATAAAGGCTACAGCTGATGCTGCGCC 1110  
QY 446 HisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGlu 465  
|||  
:::  
:::  
Db 1111 CCAGACAGACAGAGGAGATCAGCCGCTGATGAGNATGCTAGTCAACCTGGATGCC 1170  
QY 466 LeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGly 485  
|||  
:::  
:::  
Db 1171 -----TACATCCAGGAATTTGGAATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1224  
QY 486 LysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro 505  
|||  
:::  
:::  
Db 1225 CGTGTCTGCGCGCCCATCTTTCAGTATGCGCGCGGAAACCGGCGCATTTGCTACGCC 1284  
QY 506 -----ValIleArgGlyPheGlnGluLysGlnLeu---AsnValValPro 519  
|||  
:::  
:::  
Db 1285 AACCAAGTGTCTGGGACATGCGCGG-----AAGCNGTCTTACATGGGATGGAG 1335  
QY 520 GluLysGluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeu 539  
|||  
:::  
:::  
Db 1336 ATCAAGTCTGGGCCATCGCTCTTCGACCCACCAAGAA-----CAGTGTGCA 1383  
QY 540 GluGluAsnAspValLysPheTyrThrGluLeuIle-Gly----- 553  
|||  
:::  
:::  
Db 1384 GAAGAG-----GTGCTCAAGAACTTCACAGACAGCTTCGGAAGATTTCCAAAGGATGCA 1437  
QY 554 -GlyCysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIle 573  
|||  
:::  
:::  
Db 1438 GGGATGCCCATCCAGGGTCAGCATGCTTCTGCNAATACGACAGGGGCA-CACAGCGT 1496  
QY 573 eMetTyrAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGln 593  
|||  
:::  
:::  
Db 1497 G----- 1497  
QY 593 yIleGlyArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAs 613  
|||  
:::  
:::  
Db 1498 -----GAGCCCATGTTCCGGCATCTCAAGAA 1523

```
QY 613 pLysGluGlnLysValLeuMetPheIleIleIleSerLysArgGlnLeuAsnAlaTyrGl 633
Db 1524 TACCTACTCAGGACTGCATCTATTGCTCATCTGCCCGGAAGACGCCAGTGATGC 1583
QY 633 yPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGl 653
Db 1584 TGAAGTGAACGTGTGGGATACACTCTTGGGATGGCAACACACTGTGTGCAGGTGA 1643
QY 653 uThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTy 673
Db 1644 GAATGCGGTCAAGACG-----TCACCGCAGACTCTGTC 1676
QY 673 rGlnIleAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTrpse 693
Db 1677 CAACCTCTCCCTCAAGATCAATGCAACTCGGTGGCATTAACAACATCCTTA----- 1728
QY 693 rGluIleAlaGluIleSerProGluGluLysGluArgLysThrMetProLeuThrMe 713
Db 1729 -----GTCCACACAGCGCTCTGCTGTTTTCACACCGCATG---AT 1769
QY 713 tTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAl 733
Db 1770 TTTCTTGGGAGCAGATGTACACATCCCGCAGCTGGGATGGGAAGAACCATCTATCAC 1829
QY 733 aAlaValValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGl 753
Db 1830 AGCAGGTGGTAGGCAGCATGGACGCACACCCCGCATACTGTGCCACCGTGGTGTGCA 1889
QY 753 nGluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLe 773
Db 1890 G-----AlaLysPheValLys---LeuLeuArgGluPheAlaGluAsnAspAs 1910
QY 773 uGlu-----AlaLysPheValLys---LeuLeuArgGluPheAlaGluAsnAspAs 790
Db 1911 TGAAGACTTGTCTATATGTGTGGCGAGCTGTCTTATCCAGTCTTACAAGTCCACCGCTT 1970
QY 790 nArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuAr 810
Db 1971 CAAG---CCACCCCGCATCATCTTCTACCGGACGGCGTCCCGGAGGGCAGTGCCTCCA 2027
QY 810 gValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluAr 830
Db 2028 GATTTCTTCACTATGACTGTTGCCATTCCAGATGATGCATCAATCAACTG----- 2076
QY 830 gAspGlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHisAsnTh 850
Db 2077 ---GAAAAAGACTACCGCTGGGATCATATATTGTGTGTCAGAGCGCCATCACAC 2132
QY 850 rArgLeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeuThrProAl 870
Db 2133 CGCCTGTTC---TGTGCTGACAAGATGAGCGGATGGGAGAGTGGTAACATCCACG 2189
QY 870 aGluThrAspValAlaAlaValLysGlnTrpGluGluAspMetLysGluSerLy 890
Db 2190 A----- 2190
QY 890 sGluThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLy 910
Db 2191 -----GGGACCACTGTGGACACCAACATCACCCACCC 2222
QY 910 sTyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGlyThrSerArgProGl 930
Db 2223 ATTGTAGTTCGATCTTCTATCTGTGCAGCCATGCGGATCCGAGGGTACCAGCGCACATC 2282
QY 930 yHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetTh 950
Db 2283 CCATTATTATGCTCTTGGGATGACAAACCGTTTTCACGACGATGAATCCAGATCTTGAC 2342
QY 950 rTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValProVa 970
Db 2343 ATACCAGCTGTGCCACACTTACGTACGATGCACACGTTCTGTCTCTATCCACGACCTGC 2402
```

---

```
QY 970 lHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHi 990
Db 2403 CTACTATGCTCCTCTGTGGCTTTCGGGCACGA---TACCACCTGGTGCAGACGACACA 2459
QY 990 sTyrIleGlyAsp 994
Db 2460 TGACAGTGGAGAG 2472

RESULT 9
BC077863 3298 bp mRNA linear VRT 12-OCT-2004
Xenopus laevis eukaryotic translation initiation factor 2C, 1, mRNA
(cDNA clone MGC:80582 IMAGE:5440319), complete cds.
BC077863
BC077863.1 GI:50418262
MGC.
Xenopus laevis
Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus;
1 (bases 1 to 3298)
Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.
and Richardson, P.
Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative
Dev. Dyn. 225 (4), 384-391 (2002)
12454917
2 (bases 1 to 3298)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
3 (bases 1 to 3298)
Klein, S. and Gerhard, D.S.
Direct Submission
Submitted (19-JUL-2004) National Institutes of Health, Xenopus Gene
Collection (XGC), National Institute of Child Health and Human
Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
20892-7510, USA
NIH-MGC Project
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Igor Dawid
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Anura Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu,
Parvaneh Saeedi, JR Santos, Angeliqne Schnerch, Ursula Skalska,
```

Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacques Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 154 Row: c Column: 4  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

# FEATURES

```

Location/Qualifiers
1..3298
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="MGC:80582 IMAGE:5440319"
/tissue_type="Embryo, stage 31/32, Xenopus"
/clone_lib="NICHG_XGC_Emb4"
/lab_host="DH108"
/note="Vector: pCMV-SPORT6"

gene
1..3298
/genes="eif2c1-prov"
/note="synonym: MGC80582"
/db_xref="GeneID:446823"
263..2851
/genes="eif2c1-prov"
/codon_start=1
/product="eif2c1-prov protein"
/protein_id="AAH7863.1"
/db_xref="GI:50418263"
/db_xref="GeneID:446823"
/translators="MYSAGPVLVPTTTPPLMPATFKPPPPDFGTSGRITKLOA
NVPEMDIPKIEIYHYDIDIKPKCPRVNREIVHVVQHFQAIQFGRKPFVDFGRKRL
TYAMPIDKQVLEUVEVLPGEGKRIIFKVAIKWACVSLQALHDLGRPLNVPFE
TQALDVWRHLPSMYRTPVGRSFFTFASECANPLGGREHWFQGVSRPSLWOML
NIDSVATYKAQVIEFECMEQVDFKSIIEQKPLDSORVKFKIEIKGLKVEITHCG
QMKRYCNVNRVPASHOTFPLQSGGVTECTVAQYFKDRHKLVLRYPLPCLQVG
OEKHYVLPLEVCNI VAGORCIKKLTDNOSTWIRATARSAPQREISKLMSASFN
TFPWFREFGLWKMDDMTGVRVLQPPSILYGGRSKAIATPVQGVMDMKNQPHTEIE
IKWALACAPQCEVHLKTEFLQRLKISRDAGMPIQCPQCFKYAQGADSEVPMF
RLHNTYGLQVLVLTLPKGTVPYAEVRVGVLDVLMGATQCVQMKVQRTTPQTLNL
CLKINVLKAGVNNILIPQGRPPVQFQVPIFLGADVTHPPAGDGKPSIAAIVGMDAH
PNRYCAVQVQHRBIIODLSAMVRELLIOFYKSTRFKPRTIIFVRDGVSEGFQOV
LHHELLAIREACIKLEKDYOPGITFLVQKQBHRLFCCTDBNRVCKSGNIPAGTTVD
TKLTHRSFDFPLCSHAGIQGTSRPSHYHVLWDDNRPSSDELQILTYLQCHTYVRCR
SVSIPAPATYAHVLVAPRAYHLVDKEHDSAGSHTSGQNGRDRQALAKAVQVHDTL
RTMYFA"

ORIGIN
Alignment Scores:
Pred. No.: 3,56e-55 Length: 3298
Score: 567.00 Matches: 245
Percent Similarity: 40.1% Conservative: 156
Best Local Similarity: 24.5% Mismatches: 387
Query Match: 10.6% Indels: 213
DB: 11 Gaps: 39

US-10-645-746-3 (1-1020) x BC077863 (1-3298)

QY 11 GlyPheTyrArgHisSerLeuAspProGluMetLysThrLeuAlaArgProThrGlyLys 30
Db 325 GGCATATACCTTCAAGCCACCCTCCAGCGGCAGATTGTTGGTACCTCAGGGAGGACCATCAA 384
QY 31 CysAspGlyLys-PheTyrGluLysLysValLeuLeuLeuValAsnTrpPheLysPheSe 50
Db 385 GCTCCAGGCAACAGTGTGTAATGACATT----- 415
QY 50 rSerLysIleTyrAspArgLysTyrTrpGluTyrGluValLysMetThrLysGluValLe 70
Db 416 -CCCAAAATT-----GAATATATCATGATGACATTGATATA----- 451
QY 70 uAsnArgLysProGlyLysProPheProLysThrGluIleProIleProAspArgAl 90
Db 452 -----AAACCAGAGAA---TGCCCCAGACGAGTAACAGCGGAATAGTGAACACAT 501

```

```

QY 90 aLysLeuPheTrpInHisLeuArgHisGluLysGlnThrAspPheIleLeuGluAs 110
Db 502 G-----GTCCAGCAGCTTTAAAGCTCAA-----ATCTTTGGGGA 534
QY 110 pTyr-----ValPheAspGluLysAspThrValTyrSerValCysArgLeuAsnThrVa 128
Db 535 TCGAAACACAGTGTGTTGATCGAAGAAAAACCTTTATATCTGCAATGCTCTGCCAATGTC 594
QY 128 lThrSerLysMetLeuValSerGluLysValValLysLysAspSerGluLysLysAspGl 148
Db 595 AAGAGATAACACAGTG-----GAGTTAGAGTCACACTACCCGGCGAAGGAAGATCG 648
QY 148 uLysAspLeuGluLysLysIleLeuTyrThrMetIleLeuThrTyrArgLysLysPheHi 168
Db 649 T-----ATCTTCAAGTTGCAATCAAGTCGATGTAAGCTTGAAGCTTGAAGCCCTGCA 702
QY 168 sLeuAsnPheSerArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyrLysPh 188
Db 703 CGAGCTCTGCTGGTGGTCTTCCCAATGTCCTTTTGAG-----ACAGTTTCAGGC 753
QY 188 eLeuLysAsnValMetThrGln-----LysValArgTyrAlaProPheValAsnGluGl 206
Db 754 TTTAGAGCTTGTGATGAGACATTACCTTCTATGAGGTATACACCTGTC----- 802
QY 206 uIleLysValGlnPheAlaLysAsnPheValTyrAspAsnAsnSerIleLeuArgValPr 226
Db 803 -----GGTCTGCTGTTTTTACTGTCATC 825
QY 226 oGluSerPheHisAspProAsnArgPheGluGlnSerLeuGluValAlaProArgIleGl 246
Db 826 TGAAGCTCGGCAAAATCCT-----CTTGGGGGTGGTAGAGA 861
QY 246 uAlaTrpPheGlyIleTyrIleGlyIleLysGluLeuPheAspGlyGluProValLeuAs 266
Db 862 AGTTTGGTTGGTTCATCAGTCTGTACGACCTTCCCTC--TGGAAATATGATGCTCAA 918
QY 266 nPheAlaIleValAspLysLeuPheTyrAsnAlaProLysMetSerLeuLeuAspTyrLe 286
Db 919 TATCGATGTGTCTGCACAGCATTTTACAAGCA-----CAACCATGTAATGTAATTTAT 972
QY 286 uLeuLeuIleValAspProGlnSerCysAsnAspAspValArgLysAspLeuLysThrLy 306
Db 973 GTGTGAAGTCTTGGATTTTAAAGCATCGAAGACAA----- 1009
QY 306 sLeuMetAlaGlyLysMetThrIleArgGlnAlaAlaArgProArgIleArgGlnLeuLe 326
Db 1010 -----CAGAAACCTCTTACCGATTCCCAAGGGTTAAGTTTACCAAGAAAT 1056
QY 326 uGluAsnLeuLysLysCysAlaGluValTrpAspAsnGluMetSerArgLeuThrGl 346
Db 1057 TAAAGGTTTAAAGTCGAATAAAC-TCACTGTGGACAAATGAA-----ACGGA 1103
QY 346 uArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGl 366
Db 1104 A-----ATACCGAGTTTGCATGTAAACAGCGACACCGCATCATCAACATCCCTC 1157
QY 366 yLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeu-PheLysIleTyrG 386
Db 1158 TTCAGCAGGAGAGTGACAGACGGTCGAGTGTACCTGCGACACAGTACTTCAAG-----G 1211
QY 386 luGluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGlyA 406
Db 1212 ACAGACACAAACAGTAGTTCTGCGTACCTCATTTACCATGTTTACAGGTGGCGAGGAGC 1271
QY 406 lAlysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrL 426
Db 1272 AGAAACACACAGTACTCTCCCTTGAAGGTGTGCAATATATAGTCGACGCCAACAGATGCA 1331
QY 426 yAsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr---ArgLysP 445
Db 1332 AGAACTTACGGAC-----AATCAGACCTCCACTATGATACGGGCACTGCTCGATCTG 1385

```



```
QY 445 roHieaspTyrLysGluAenThrLeuLysMetLeuLysGluLeuAepPheSerSerGluG 465
Db 1386 CTCACAGATCCGAGGAAGAGATCAGTAAATTAATGCGAAGTGCAGTTTAACTACTGATC 1445
QY 465 luLeuAenPheValGluArgPheGlyLeuCySerLysLeuGlnMetIleGluCyProG 485
Db 1446 CC-----TTTGTGCGAGAGTTTGGTATTAATGGTGAAGATCAGATGACCGATGTCACAG 1499
QY 485 llyValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrP 505
Db 1500 GCAGAGTCCTCAACCTCTTCAATCTTTATGAGGAGGAGTAAAGCAATTGCACGC 1559
QY 505 roVal-----lleArgGlyPheGlnGluLysGlnLeuAenVal---ValProGluLysG 522
Db 1560 CTGTTCCAGGTGCTGGGATATGAGAAACAACAGTTCACACCCGCGCATTTGAAATAAAGG 1619
QY 522 luLeuCyCysAlaValPheValValAsnGluThrAlaGlyAsnProCysLeuGluGluA 542
Db 1620 TGTGGGCCATCGCTGCTTTGCCCCACAA-----CGACAGTGTACTGAAGTTC 1667
QY 542 enAspValLysPheTyrThrGluLeuIle-Gly-----GlyCys 555
Db 1668 AC-----CTGAAGACATTTACAGAACAGTTGAGGAAGATTCTCGTGATGCCGGGATGC 1721
QY 556 LysPheArgGlyIleArgIleGlyAlaAsnGluAenArgGlyAlaGlnSerIle----- 573
Db 1722 CCATCCAGGGGAGCCCATGTTCTGCAGATAT-GCCAGGAGCAGATAGTGTGAGCCA 1780
QY 574 MetTyrAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAenThrGly 593
Db 1781 ATGTTCCAGGCATCTGAAGATACATAT-----ACAGGG 1813
QY 594 lleGlyArgPheGluIleAlaThrGluAlaLysAsnMetPheGluArgLeuProAsp 613
Db 1814 CTA----- 1816
QY 614 LysGluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAenAlaTyrGly 633
Db 1817 -----CAACTGTTGTGTAATCTGCTCGGGAAACCCCTGCTATGGG 1861
QY 634 PheValLysHieTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGlu 653
Db 1862 GAAGTGAACCGTGTGGGCACACTGTTCTGGGAATGGCTACTCAGTGTGTTCCAGATGAAA 1921
QY 654 ThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyr 673
Db 1922 AATGTG-----CAGAGACCACACACCAAACTGTGCT 1954
QY 674 GlnIleAlaLeuLysIleAsnAlaLysLeuGlyIleAsnGlnGluLeuAepTrpSer 693
Db 1955 AACCTGTGCTTAAATAATAATGTAATAATTAGAGGGCTCAATAACATTTTG----- 2005
QY 694 GluIleAlaGluIleSerProGluGluLysGluArgArgLysThrMetProLeuThrMet 713
Db 2006 -----TTGCCCCAGGCAGACACCTCTGCTGTTTCAACAGGCAGTC---ATA 2047
QY 714 TyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIleAla 733
Db 2048 TTTCTGGAGCTGATGTTTACCACCCACCTGAGGGGACGGTAAAGACCTTCGATGCT 2107
QY 734 AlaValValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValThrGln 753
Db 2108 GGGGTTGTAGTATGATGATGACACCCCAACCGATATGTGCCACTGTGGGGTGCAG 2167
QY 754 GluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeu 773
Db 2168 CAG-----CACCAGCAAGAAATCATTCAGGATTTG 2197
QY 774 GluAlaLysPheValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaPro 793
Db 2198 TCTGCATGCTTGGGAGCTGCTTATCCAGTTTCTACAAGTCTACTGCTGTCAAG---CCC 2254
QY 794 AlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHis 813
```

```
Db 2255 ACNAGAATCACTTTTACAGAGATGGTGTCTCAGGAAGGCGAGTTCACACAGTTCCTCCAT 2314
QY 814 AspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGlu 833
Db 2315 CATGAGCTTCGGCCATTAGGAAGCCCTGTATTAAACTG-----GAGAAG 2359
QY 834 AspProGluProLysTyrThrPheIleValIleGlnLysAenThrArgLeuLeu 853
Db 2360 GATTACCAGCCAGGAATTCATTTATTGTAGTGCAAAAGACATCACACGGCGCTATTTC 2419
QY 854 ArgAspMetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAsp 873
Db 2420 ---TGTCAGACCCGAAATGAGAGGCTAGGAAAGTGGAAATATACCAGCT----- 2467
QY 874 ValAlaValAlaAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGly 893
Db 2467 ----- 2467
QY 894 lleValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPhe 913
Db 2468 -----GGTACCACCTGTAGACACAAAGATTACCACCCCTTCAGAGTTT 2509
QY 914 AspPhePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThr 933
Db 2510 GATTTTACCTGTGCAGTCAGCTGCTGATTTCAGGGAACAAGCAGCGTCACACTATCAT 2569
QY 934 ValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeu 953
Db 2570 GTCTGTGGGATGACAATCGCTTTTCATCTGATGATTTGCAGATCCTCATTCAGCTC 2629
QY 954 AlaPheLeuSerAlaArgCysArgLysProIleSerLeuProValHisTyrAla 973
Db 2630 TGCCACACTTATGTTGTTGCACCTGCTGCTGTCTATCTCCCTGCTCCAGCATATATGCT 2689
QY 974 HisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990
Db 2690 CACCTGTGGCATTCAGAGCAAG--TATCATTTTGGTGGATAAAGAGCAT 2737

RESULT 10
AX575501 3580 bp DNA linear PAT 07-JAN-2003
LOCUS Sequence 17 from Patent WO02072630.
DEFINITION AX575501
ACCESSION AX575501
VERSION AX575501.1 GI:27552093
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1
AUTHORS Thornton, M., Hafalia, A.J., Lu, D.A., Arvizu, C., Swarnakar, A., Lu, Y.,
Warren, B.A., Baughn, M.R., Tang, Y.T., Lee, E.A., Yao, M.G.,
Ramkumar, J., Khan, F.A., Gandhi, A.R., Ding, L., Yue, H., Gietzen, K.J.,
Walia, N.K., Thangavelu, K., Elliot, V.S. and Marquis, J.P.
TITLE Nucleic acid-associated proteins
JOURNAL Patent: WO 02072630-A 17 19-SEP-2002;
Incyte Genomics, Inc. (US)
FEATURES
source
1..3580
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 1725129CB1"

ORIGIN
Alignment Scores: 5.26e-55 Length: 3580
Pred. No.: 239
Score: 566.00 Matches: 239
Percent Similarity: 40.0% Conservative: 159
Best Local Similarity: 24.0% Mismatches: 385
Query Match: 10.6%
```



[illegible][illegible]

```
Db 1781 -----CTGCCCGGAGGCGCGCGGTGTTCACGAGCGCGTC---ATCTTTCT 1827
Qy 715 lgllyleAapValThHisProThSerTyrSerGlylleAapTyrSerIleAalaVa 735
Db 1828 GGGAGCAGACGTCACCTACCCCGCGGGATGGGAAGAAGCCCTCCATTGCGCGCT 1887
Qy 735 lvalAlaSerIleAenProGlyGlyThrIleTyrArgAenMetIleValThrGlnGlu 755
Db 1888 GTGGGCACATGAGGACCCACCCCATCGCTACTCGCCACCGTGGCGGTGCAGCAG-- 1945
Qy 755 uCySArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIleLeuGluAl 775
Db 1946 -----CACCGGCAGGAGATCATACAAGACCTGCGCCG 1977
Qy 775 alySpheVallyLeuLeuArgGluPheAlaGluAenAenAapAenArgAlaProAlaHi 795
Db 1978 CATGTCGCGAGCTCTCATCCAGTCTTACAAGTCCACGCGCTCAAG---CCCACCG 2034
Qy 795 stIleValValTyrArgAapGlyValSerAapSerGluMetLeuArgValSerHiAaspGl 815
Db 2035 CATCATCTTCTACCGCAGCGGTCTCTGAGCCAGTTCACGACGTTCTCCACACCA 2094
Qy 815 uLeuArgSerLeuLysSerGluVallyLysGlnPheMetSerGluArgAapGlyGluAapPr 835
Db 2095 GTTGTGCGCATCCGTCGAGCGCTGTATCAAGCTA-----GAAAAAGACTA 2139
Qy 835 oGluProLysTyrThrPheIleValleGlnLysArgHisAenThrArgLeuArgAr 855
Db 2140 CCAGCCCGGAGTACCTTCATCGTGTGTCAGAGAGGACACACCGCGCTTTC---TG 2196
Qy 855 gMetGluLysAapLysProValValAenLysAapLeuThrProAlaGluThrAspValAl 875
Db 2197 CACTGACAAGACGAGCGGTGGGNAAGTGGAAACATCCAGCA----- 2242
Qy 875 aValAlaValLysGlnTrpGluLysMetLysGluSerLysGluThrGlyIleVa 895
Db 2242 ----- 2242
Qy 895 lAenProSerSerGlyThrThValAspLysIleuLeuValSerLysTyrLysPheAapPh 915
Db 2243 -----GGCACACACTGTGGACACGAAATCACCCACCCACCGAGTTGCAGT 2289
Qy 915 ePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMe 935
Db 2290 CTACCTGTGTATCAGCTGGATCCAGGGACAGAGCCCTTCCACTATCAGCTCT 2349
Qy 935 tTyrAapAapLysGlyMetSerGlnAapGluValTyrLysMetThrTyrGlyLeuAlaPh 955
Db 2350 CTGGGACGACAATCGTTTCTCTCTGATGCTGACATCCTAAACCTACGCTGTGTCA 2409
Qy 955 eLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHisle 975
Db 2410 CACCTACGTCGCTGCACACGTCGCTGTCATCCAGCGCCAGCATACTACGCTCACCT 2469
Qy 975 uSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990
Db 2470 GTGGCTTCGCGGCGAGS---TACCACCTGTGTGNTAAGAACAT 2512

RESULT 11
BC075263
LOCUS
DEFINITION
Xenopus tropicalis MGC88879 protein, mRNA (cdna clone MGC:88879
IMAGE:6988277), complete cds.
ACCESSION
BC075263
VERSION
BC075263.1 GI:49522951
KEYWORDS
MGC.
SOURCE
Xenopus tropicalis (Silurana tropicalis)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 3687)
AUTHORS
Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
```

Collins FS, Wagner L, Shennan CM, Schuler GD, Altschul SF, Zeeberg B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T, Max SI, Wang J, Hsien F, Diatchenko L, Maruina K, Farmer AA, Rubin GM, Hong L, Stapleton M, Soares MB, Bonaldo MP, Casavant TL, Scheetz TB, Brownstein MJ, Ustin TB, Toshiyuki S, Carninci P, Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy SJ, Bosak SA, McSwan PJ, McKernan KJ, Malek JA, Gunaratne PH, Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulvik SW, Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J, Helton E, Kettaman M, Madan A, Rodrigues S, Sanchez A, Whiting M, Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW, Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J, Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U, Smailus DE, Schnerch A, Schein JE, Jones SJ and Marra MA. Mammalian Gene Collection Program Team  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 12477932  
2 (bases 1 to 3687)  
Klein,S. and Gerhard,D.S.  
Direct Submission  
Submitted (28-JUN-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA  
NIH-MGC Project  
Contact: XGC help desk  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Rob Granger, University of Virginia  
cDNA Library Preparation: Open Biosystems  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca  
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Anara Maason, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prähbu, Farvaneh Saeedi, JR Santos, Angeliq Schnerch, Ursula Skalska, Duane Smalls, Jeff Stott, Miranda Tsai, George Yang, Jacque Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 171 Row: a Column: 14  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 52345659.

FEATURES  
source

1. 3687  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="MGC:88879 IMAGE:6988277"  
/tissue\_type="Whole body, male, 10 months old, strain F6, normal"  
/clone\_lib="NICHD\_XGC\_SwblN"  
/lab\_host="DH10B"  
/note="Vector: pExpress1"  
1. 3687  
/gene="MGC88879"  
/db\_xref="GeneID:448205"  
239..2854  
/gene="MGC88879"  
/codon\_start=1  
/product="MGC88879 protein"  
/protein\_id="AAH75263.1"  
/db\_xref="GI:49522952"  
/db\_xref="GeneID:448205"  
/translation="MYAGAPVLVPTTPPLPMAYTFKPPRPDQCTSGRTIKLOA  
NFFENDIPKIEIHYEIDIKPEKPRRINVEHNVQHFKAQIFGDRAPFGDRKRL  
YTAMPLPTARDKQVELVTLFPEGKDRIFKVAIKWMAVCVSLQALHDALSLVSPVFE





## Alignment Scores:

Pred. No.: 4,45e-55 Length: 2657  
 Score: 565.00 Matches: 231  
 Percent Similarity: 39.2% Conservative: 159  
 Best Local Similarity: 23.2% Mismatches: 355  
 Query Match: 10.6% Indels: 250  
 DB: 6 Gaps: 40

US-10-645-746-3 (1-1020) x AB081472 (1-2657)

Qy 38 LysLysValLeuLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGlu 57  
 Db 136 AGAACATCAACATACAGGCCAATTTCTTTGAAATGGACATCCCAAAATGGACATCAT 195  
 Qy 58 TyrTrpGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysPro 77  
 Db 196 CACTATCAATTTGGACATCAACCTTGAGAAACGC----- 228  
 Qy 78 PheProLysLysThrGluIleProIleProAspArgAlaLysLeuPheTrpGlnHisLeu 97  
 Db 229 ---CCTCGGAGAGTAACAGGAAATTTGTGAGCAGCATG-----GTC 267  
 Qy 98 ArgHisGluLysLysGlnThrAspPheIleLeuGluAspTyr-----ValPheAspGlu 115  
 Db 268 CAGCACATTTAAACCTCAG-----ATCTCGGGGACCGGAAGCCAGTGTGTGATGGA 318  
 Qy 116 LysAspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSer 135  
 Db 319 AGGAAGAATCTGTACACAGCAATGCCCTTCCGATCGGAGGGACAAAGTG----- 369  
 Qy 136 GluLysValValLysLysAspSerGluLysLysAspLysLysLeuLysLysIle 155  
 Db 370 GAGCTGAGGTACAGCTCCCGGAGAGGAAGCAAGAT---CGCATCTTTAAAGGTATCCATC 426  
 Qy 156 LeuTyrThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSerArgGluAsn 175  
 Db 427 AAGTGGGTGTCGTCGTGACCTG---CAGGCGTTACACGATGCATCTTCGGGGCGGCTG 483  
 Qy 176 ProGluLysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnValMetThrGln 195  
 Db 484 CCCAGCGTCCCTTCGAG-----ACGATCCAGGCCCTCGGACGTTGTATGAGGCAC 534  
 Qy 196 -----LysValArgTyrAlaProPheValAsnGluLuleLysValGlnPheAlaLys 213  
 Db 535 TTACCATCCATGAGGTACACCCCTGTT----- 561  
 Qy 214 AsnPheValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspProAsn 233  
 Db 562 -----GGCCGTTCTTCTTCATCGCATCTGAGGCTGTTCACACCTT--- 603  
 Qy 234 ArgPheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyrIle 253  
 Db 604 -----CTGGGTGGGGGACAGAGTGTGTTGGCTTCCATCAG 642  
 Qy 254 GlyIleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeu 273  
 Db 643 TCGTCCGCTCTCTCT---TGGAAATGATGCTGAATATGATGATGATGCGCAACAGCG 699  
 Qy 274 PheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuLeuIleValAspProGln 293  
 Db 700 TTTTACAAGGCA-----CAGCCAGTGTGAGTGTGTTGTGCAAGTTTGGATTATAA 753  
 Qy 294 SerCysAsnAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThr 313  
 Db 754 AGTATTCAAGAACAA-----CAAAAAACCT 777  
 Qy 314 IleArgGluAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLys--- 332  
 Db 778 CTGACAGATTTCCCAAGGGTAAAGTTTACCAAGAAATCAAAAGTCTTAAAGGTGGAGATA 837  
 Qy 333 -----CysAlaGluValTrpAspAsn-----GluMetSerArgLeuThrGlu 346

Db 838 ACGCACTGTGTGTCAGATGAAGAGAACTACCGTGTCTGCAATGTGACCCGGCGGCTGCC 897  
 Qy 347 ArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThrGly 366  
 Db 898 AGTCACCAACAGTTTC---CCACTGCAGCAGGAGAGTGGGACAGACAGTGGAGTGATACA--- 951  
 Qy 367 LysSerAspArgGlyArgAsnAlaLysLysTyrAspThrThrLeuPheLysIleTyrGlu 386  
 Db 952 -----GTGGCCCACTGATC-----TTCAAG-----GAC 972  
 Qy 387 GluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGlyAla 406  
 Db 973 AGGCACAAAGCTGTTCTGCGCTACCCCACTCCCGTGTGTACAAAGTGGGACAGAGCG 1032  
 Qy 407 LysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyrLys 426  
 Db 1033 AAACACACCTACCTCTCTTTGGAGTCTGTAACATAGTGTGTCGACAGAGATGTATAAA 1092  
 Qy 427 AsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr-----ArgLysPro 445  
 Db 1093 AAATTAACAGAC-----GATCAGACCTCAACCATGATCAGACCACTGCCAGTGCAGCA 1146  
 Qy 446 HisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGlu 465  
 Db 1147 CCTGATCGCCAAAGAGGATCAGCAAACTGATGGAAGTGGAAATTTCAATACAGACCCA 1206  
 Qy 466 LeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysProGly 485  
 Db 1207 -----TATGTTCTGTAATTTGGAATCATCGTGAAGATGAGATGACAGATGTGACTGGG 1260  
 Qy 486 LysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro 505  
 Db 1261 CGGGTTCTGACGCGGCCCTCATCTCTACGGGGGACGAAACAAGCAATTTGCCACCCCT 1320  
 Qy 506 Val-----IleArgGlyPheGlnGluLysGluLeuAsnVal----- 517  
 Db 1321 GTCCAGGGTGTGGGACATCGCAACAACAGCATTCACAGGGCATCGAGATCAGAGTG 1380  
 Qy 518 -----ValProGluLysGluLeuCysYsAlaValPheValValAsn 531  
 Db 1381 TGGGCCATCGCCTGCTTGTCTCCCGAGCGCAG---TGATACAGAGTCCATCTCAAGTCC 1437  
 Qy 532 GluThr-----AlaGlyAsnPro 537  
 Db 1438 TTCACAGACCACTGAGGAAGATCTCGAGAGATGTGGGATGCCGATCCAGGGTGCAGCCC 1497  
 Qy 538 CysLeuGluGluAsnAspValValLysPheTyrThrGluLeuIleGlyGlyCysLysPhe 557  
 Db 1498 TGTTC-----TGTAATATC 1512  
 Qy 558 ArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSerIle-----MetTyr 575  
 Db 1513 GCA-----CAGGGTGCAGACAGTGTGGAGGCCATGTTTC 1545  
 Qy 576 AspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnThrGlyIleGly 595  
 Db 1546 CGACACCTCAAGAACACATACGCT----- 1569  
 Qy 596 ArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuProAspLysGlu 615  
 Db 1569 ----- 1569  
 Qy 616 GlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTyrGlyPheVal 635  
 Db 1570 ---GGCCTCCAGCTGTGTGTGTCTCTCCCTGGCAAACTCCTGTGTATGCGGAAGTC 1626  
 Qy 636 LysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisIleThrSerGluThrVal 655  
 Db 1627 AAGCGTGTGGGAGACACAGTGTGGGGATGGCCACACAGTGTGCTCCAGATGAAGAAGCTG 1686  
 Qy 656 ThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePheTyrGlnIle 675  
 Db 1687 -----CAGAGGACCCAGCCACAGACCCCTATCCAAATCTC 1719

```
QY 676 AlaLeuValLeuAlaAsnAlaLysLeuGlyGlyLeuAlaGlnGlnLeuAspTrpSerGluLeu 695
DB 1720 TGCTTAAAGATCAATGTCACAACTGGGAGCGTCACCAACATCCTG-----1764
QY 696 AlaGluLeuSerProGluGluLysGluAlaGlySerThrMetProLeuThrMetTyrVal 715
DB 1765 -----CTGCCACAGGCGAGCTCCAGTGTTCACGACACCTGTC---ATCTCTCG 1812
QY 716 GlyLeuAspValThrHisProThrSerTyrSerGlyLeuAspTyrSerLeuAlaVal 735
DB 1813 CGAGCGCATGTCAACCCACCCAGCTGGGAGTGGGAAGAAACCTTCTATTGCTGCGGTT 1872
QY 736 ValAlaSerLeuAsnProGlyThrLeuTyrArgAsnMetLeuValThrGlnGlu 755
DB 1873 GTGGCAGCATGGACGACACACCCCAACCGCTACTGTGCCACCGTGGCGTGCAGCAGC---1929
QY 756 CysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspLeuGluAla 775
DB 1930 -----CACCCACAGGAGATCATCCAGGACCTGGCTGCC 1962
QY 776 LysPheValLysLeuLeuArgGluPheAlaGluAsnAsnAspAsnArgAlaProLahis 795
DB 1963 ATGTTACGAGAGTGTCTCATTCAGTCTTCAAGTCCACCGCTTCAAG---CCACCCGC 2019
QY 796 IleValValTyrArgAspGlyValSerAspSerGluMetLeuArgValSerHisAspGlu 815
DB 2020 ATCATCTTACCGCATGCGCTCTCCGAGGCGCAATCCAGCAGGTTCACACCATGAG 2079
QY 816 LeuArgSerLeuLysSerGluValLysGlnPheMetSerGluArgAspGlyGluAspPro 835
DB 2080 CTCTCGCCATCAGAGCGCTGCATCAGCTG-----CAGAGGACTAT 2124
QY 836 GluProLysTyrThrPheIleValIleGlnLysArgHisAsnThrArgLeuLeuArgArg 855
DB 2125 CAGCCAGGAATCAGTTCATCGTGTGCAGAGCGGCCACACACACCGCTCTTC---TGC 2181
QY 856 MetGluLysAspLysProValValAsnLysAspLeuThrProAlaGluThrAspValAla 875
DB 2182 ACAGACAAAATGAGCGGCTGGGAAGAGTGGGAACATTCCTCCGCA-----2226
QY 876 ValAlaAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyLeuVal 895
DB 2226 -----2226
QY 896 AsnProSerSerGlyThrThrValAspLysLeuIleValSerLysTyrLysPheAspPhe 915
DB 2227 -----GGCACAACCGTGGACAGCAAGATCACCCACCCCACTGAGTTGACTTC 2274
QY 916 PheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisTyrThrValMet 935
DB 2275 TACCTGTGCAGTCACCGCGCATCCAGGGGACAGCGCGTCTTCCCACTACCACTGCTT 2334
QY 936 TyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAlaPhe 955
DB 2335 TGGGATGACAATCGTTTTTCTCTGATGAGCTGCAGATCTTGACCTACCACTGCTGCAC 2394
QY 956 LeuSerAlaArgCysArgLysProLysLeuProValProValHisTyrAlaHisLeu 975
DB 2395 ACCTACGTGCTGCACAGCTCTGTGTCAATACCGCGCCAGCTTACTATGCTCACCTG 2454
QY 976 SerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990
DB 2455 GTGGCCTCCGAGCCAGG---TACCACCTGGTGGATAAAGAACAT 2496
RESULT 13
AF005355 LOCUS
DEFINITION Oryctolagus cuniculus translation initiation factor eIF2C mRNA,
complete cds.
ACCESSION AF005355
VERSION AF005355.1 GI:3253158
KEYWORDS
```

## SOURCE

## ORGANISM

Oryctolagus cuniculus (rabbit)

Oryctolagus cuniculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha;  
Leporidae; Oryctolagus.

1 (bases 1 to 3599)

Zou, C., Zhang, Z., Wu, S. and Osterman, J.C.

Molecular cloning and characterization of a rabbit eIF2C protein

Gene 211 (2), 187-194 (1998)

2 (bases 1 to 3599)

Zou, C., Zhang, Z., Wu, S. and Osterman, J.C.

Direct Submission

Submitted (23-MAY-1997) Chemistry, University of Nebraska, 760

Hamilton Hall, Lincoln, NE 68588, USA

Location/Qualifiers

1..3599

/organism="Oryctolagus cuniculus"

/mol\_type="mRNA"

/db\_xref="taxon:9986"

/tissue\_type="liver"

82..2523

/codon\_start=1

/product="translation initiation factor eIF2C"

/protein\_id="AAC24323.1"

/db\_xref="GI:3253159"

/translation="MDIPKIDYHYELDIKEKCPRRVNRIVEHMQHFAKQIFGDR  
KPEVDGRKNTYAMPPIGRIBKYLEVTLPGEGKDIRIFKVISIKWVSCVSLQALHDALS  
GRLPSPVFPETIQALDVMRHLPSMRYPVGRVSFFTAEGCSNPLGGREVWFGHQSV  
RPSWMMMLNIDVSATYQAQVIEFCEVLDFKSIIEQQKPLTDSRVKTKPKIKG  
KLVEITHCGQMKRYVNCVNRPPASHOTFPLOQSGQGTVECTVATQVFKDRHKLVLRY  
PLHPCLOVQSQKHTYLPLEVCNIVAGORCIKKLTNDOTSTMIRATASAPORQEIS  
KLMSASNTDPYREFGIMVKDEMTDVTGVLPPSLYGRNKAIATPVQGVMDMR  
NKQHTGLEIKWALACAPAPQCQTEVHLKSFTEGRKISRDAQNPQIQGPCFKYAQ  
GADSVGMRHLNATYAGLVVLPGKTPVYAKRVGDTVLGMATQCVQMKNVQR  
TTPQTLNKLKINLVGVNNILLPQGRPVFPVQOPVIFLGADVTHPAGDGGKPSIA  
AVGSMADPNRYCATVRVQOHROEIIQDLAAVRELLIQFYKSTRFKPTRIFVRDG  
VSEGOQOVLHHELLAIIEACIKLEKDYQPGITFIVOKRHHTRFLPCTDKNERVQSG  
NIPAGTVDTKLTPTREFYLCSHNGIQTSRPSHVLMWDDNPFSSDELQILTYQL  
CHTVRCRSVSIIPAPATYAHLVAFRARIHLVDKEDHSAEGSHTSQSGNDRHQALAK  
AVQHQDTLRTMYFA"

## ORIGIN

## Alignment Scores:

Pred. No.: 6,956-55 Length: 3599

Score: 565.00 Matches: 235

Percent Similarity: 40.2% Conservative: 158

Best Local Similarity: 24.1% Mismatches: 371

Query Match: 10.6% Indels: 214

DB: 14 Gaps: 38

US-10-645-746-3 (1-1020) x AF005355 (1-3599)

```
QY 38 LysLysValLeuLeuValAsnTrpPheLysPheSerSerLysLeuTyrAspArgGlu 57
DB 49 AGAACATCAAGTTCAGCGCCCAATTTCTTCGAGATGGACATCCCAAAATTCACATCTAT 108
QY 58 TyrTyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysPro 77
DB 109 CACTATGAATGGATATCAAGCCGAGAA-----138
QY 78 PheProLysLysThrGluLeuProLysPheProAspArgAlaLysLeuPheTrpGlnHisLeu 97
DB 139 TGCCCCAGAGAGTGAACAGGAGATAGTAGACACATG-----GTG 180
QY 98 ArgHisGluLysLysGlnThrAspPheLeuGluAspTyr-----ValPheAspGlu 115
DB 181 CAGCACTTTAAAGCCAG-----ATCTTCGGCGACCGGAAGCCGGTGTTCGACGGG 231
QY 116 LysAspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSer 135
DB 232 AGGAAGAACCTGTATACACGGCGATGCCCTGCCCCATCGGAGGAGGAAGGTG-----282
```



Qy	136	GluLysValValLysLysAspSerGluLysLysAspGluLysAspLeuGluLysLysIle	151
Db	283	GAGCTGGAGGTCAAGTTCGGCGAAGCAAGGACCGC-----ATCTTCAAGGTGTCC	336
Qy	156	LeuTyThrMetIleLeuThrTyArgLysLysPheHisLeuAsnPheSerArgGluAsn	175
Db	337	ATCAAAGTGCGTGTGGTGCAGCGTGCAGCGCGCTGCACGACGCATTTTCGGCGGCGTGG	396
Qy	176	ProGluLysAspGluAlaAsnArgSerTyLysPheLeuLysAsnValMetThrGln	195
Db	397	CCCAGCGTCCCGTTCGAG-----ACGATCCAGGCCCTGGAGCTGGTCATGAGGCAT	447
Qy	196	-----LysValArgTyThrAlaProPheValAsnGluLysValGlnPheAlaLys	213
Db	448	TTGCCATCATGAGTACACGCCTGTG-----	474
Qy	214	AsnPheValTyThrAspAsnAsnSerIleLeuArgValProGluSerPheHisnapProAsn	233
Db	475	-----GGCGCGTCTCTTCTCACCGCGTCCGAGGCGCTGCTCCAACCT---	516
Qy	234	ArgPheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheChlyIleTyrlle	253
Db	517	-----CTTCGGGGGCCCGGAAGTGTGTCTCGCTTCGCCTCCACGAC	555
Qy	254	GlyIleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLysLeu	273
Db	556	TCTGTCCGACCTTCTCTT---TGGAAATATGATCTGTAATTATTGACGTGTCACGACGCACA	612
Qy	274	PheTyrrAsnAlaProLysMetSerLeuLeuAspTyThrLeuLeuLysValAspProGln	293
Db	613	TTCTACAAGGCG-----CAGCCAGTAAATCGAGTTTGTGTGAAGTTCGTGATTTTAAA	666
Qy	294	SerCyAsnAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMetThr	313
Db	667	AGCATTGAAGAACA-----CAGAACCT	690
Qy	314	IleArgGlnAlaAlaArgProArgIleArgGlnLeuGluAsnLeuLysLeus---	332
Db	691	CTGACAGATTCCTCAAGAGGTCAAGTTTACCAGAAATCAAAAGGTCTCAAGGTGGAGATA	750
Qy	333	-----CysAlaGluValTrpAspAsn-----GluMetSerArgLeuThrGlu	346
Db	751	ACGCACGTGGCGAGATAAGAGGAAGTACCGCGCTCTGCACAGTCAACCCGCGCGCCGCC	810
Qy	347	ArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyThrLysValThrGly	366
Db	811	AGCCACACGACTTC---CCTCTGCAGCAGAGAGCGCGGAGACGCTGGAGTGCAGC---	864
Qy	367	LysSerAspArgGlyArgAsnAlaLysLysTyThrAspThrThrLeuPheLysIleTyThrGlu	386
Db	865	-----GTGGCCCCAGTAC-----TTCAAG-----GAC	885
Qy	387	GluAsnLysLysPheIleGluPheProHisLeuProLeuValLysValLysSerGlyAla	406
Db	886	AGGCACACAGCTGGTTCGTGCGTCACTCCCAACCTCCCGTCTGCAAGTGGGACAGAGCAG	945
Qy	407	LysGluTyThrAlaValProMetGluHisLeuGluValHisGluLysProGlnArgTyThrLys	426
Db	946	AAGCACACTACCTTCCCTCGAGGTCTGCACCATATGTCGCGGCGGACAGAGATGTATAAAA	1005
Qy	427	AsnArgIleAspLeuValMetGlnAspLysPheLeuLysArgAlaThr---ArgLysPro	445
Db	1006	AAATTAACGGAC-----AATCAAACTCAACTATGATCATCAGCGACCGCCAGCTCTCGG	1059
Qy	446	HisAspTyThrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGluGlu	465
Db	1060	CCCAGTCGGCAAGAGAGATCAGCAAGTTGATGCGAAGCGCGAGCTTCAACACAGATCCG	1119
Qy	466	LeuAsnPheValGluArgPheGlyLeuCySerLysLeuGlnMetIleGluCyserProGly	485
Db	1120	-----TACGTCCGCGAGTTCGGAATCATGTGTGAAGACGAGATGACGCGCTGACGCGGG	1173
Qy	486	LysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThrPro	505

[illegible]



Db 2034 CTACACCGGGGATCACCTTCATCGTGTGCAGAGCGGCACACACCGGCTCTTC--- 2091  
 Qy 854 gArgMetGluLysAspLysProValValAlaLysAspLeuThrProAlaGluThrAspVa 874  
 Db 2092 -TGCACGGACAGACAGCGGGTCCGCAAGAGCGGGAACATTCACAGC----- 2139  
 Qy 874 lAlaValAlaAlaValLysGlnTrpGluGluAspMetLysGluSerLysGluThrGlyI 894  
 Db 2139 ----- 2139  
 Qy 894 eValAsnProSerSerGlyThrThrValAspLysLeuIleValSerLysThrLysPheAs 914  
 Db 2140 -----GSCACGACCGTGCACACGAGATCACACACCGCAGGTTTGA 2183  
 Qy 914 pPhePheLeuAlaSerHisGlyValLeuGlyThrSerArgProGlyHisThrVa 934  
 Db 2184 CTTTACCTGTGTGAGCCAGCGCGCATCCAGGGGACAGCGGCCCTCCCACTACCACGT 2243  
 Qy 934 lMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMetThrTyrGlyLeuAl 954  
 Db 2244 CTTGTGGATGACATCGCTTCTTCCGATGAGCTACAGATCCTGACCTACAGCTGTG 2303  
 Qy 954 aPheLeuSerAlaArgCysArgLysProIleSerLeuProValProValHisTyrAlaHi 974  
 Db 2304 TCACACCTAGTGTGTGAGCGGCTCGTCTCTATCCCGCGCGGCTTACTAGCTCA 2363  
 Qy 974 sLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGluHis 990  
 Db 2364 CTTGTGTGCTTCCGCGCCAGG---TACCACCTGTGTGGAAGAAGACAC 2409

RESULT 14

BD156502 3996 bp DNA linear PAT 17-JAN-2003  
 LOCUS  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION BD156502  
 VERSION BD156502.1 GI:27862260  
 KEYWORDS JP 2002191363-A/11345.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Ota T, Isogai T., Nishikawa T., Hayashi K., Saito K., Yamamoto J., Ishii S., Sugiyama T., Wakamatsu A., Nagai K. and Otsuki T.  
 TITLE Primer for synthesizing full-length cDNA and use thereof  
 JOURNAL Patent: JP 2002191363-A 11345 09-JUL-2002;  
 COMMENT HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/11345  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC 10,  
 PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key

FT CDS Location/Qualifiers  
 (297)..(2642).  
 1..3996  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 8.11e-55 Length: 3996

Score: 565.00 Matches: 230

Percent Similarity: 40.6% Conservative: 170  
 Best Local Similarity: 23.3% Mismatches: 363  
 Query Match: 10.6% Indels: 224  
 DB: 2 Gaps: 37  
 US-10-645-746-3 (1-1020) x BD156502 (1-3996)  
 Qy 38 LysLysValLeuLeuLeuValAsnTrpPheLysPheSerSerLysIleTyrAspArgGlu 57  
 Db 171 AAACCAATCAAGCTCTGGCCAAATTACTTTAGGTGGACATCCCTTAAGATCGACGTGTAC 230  
 Qy 58 TyrTyrGluTyrGluValLysMetThrLysGluValLeuAsnArgLysProGlyLysPro 77  
 Db 231 CACTACAGGTGACATCAAGCCGATAG----- 260  
 Qy 78 PheProLysLysThrGluLeuProIleProAspArgAlaLysLeuPheTrpGlnHisLeu 97  
 Db 261 TGTCCCGTAGAGTCAACCGGAAGTGTGGAA-----TACATGGTC 302  
 Qy 98 ArgHisGluLysGlnThrAspPheIleLeuGluAspTyr-----ValPheAspGlu 115  
 Db 303 CAGCATTTCAAGCCTCAG-----ATCTTTGGTGATCGCAAGCCTGTGTATGAGA 353  
 Qy 116 LysAspThrValTyrSerValCysArgLeuAsnThrValThrSerLysMetLeuValSer 135  
 Db 354 AAGAAGAACATTAC-----ACTGTCACAGCACTGCCCATTTGGCAAC 395  
 Qy 136 GluLysVal-----ValLysLysAspSerGluLysLysAspGluLysAspLeuGlu 152  
 Db 396 GAACGGGTGCACTTTGAGGTGACAAATCCCTGGGGAAGGAAGGAT---CGAATCTTTAAG 452  
 Qy 153 LysLysIleLeuTyrThrMetIleLeuThrTyrArgLysLysPheHisLeuAsnPheSer 172  
 Db 453 GTCTCCATCAAGTGTGCTAGCCATTGTGAGCTGGCGAATGCTGCATGAGGCCCTGTGTCAGC 512  
 Qy 173 ArgGluAsnProGluLysAspGluGluAlaAsnArgSerTyrLysPheLeuLysAsnVal 192  
 Db 513 GCCCAGATCCCTGTTCCCTTGAGTCTGTGCAAGCCCTGGATGTGGCCATGAGGCACCTG 572  
 Qy 193 MetThrGlnLysValArgTyrAlaProPheValAsnGluGluLysValGlnPheAla 212  
 Db 573 GCATCC-----ATGAGGTACACCCCTGTG----- 596  
 Qy 213 LysAsnPheValTyrAspAsnAsnSerIleLeuArgValProGluSerPheHisAspPro 232  
 Db 597 -----GGCCGCTCTCTTCTCTACCCCTGAGGCTACTACCAACCG 638  
 Qy 233 AsnArgPheGluGlnSerLeuGluValAlaProArgIleGluAlaTrpPheGlyIleTyr 252  
 Db 639 -----CTGGGGGTGGCGCGAGGTCTGTGCTTGGCTTTCAC 674  
 Qy 253 IleGlyIleLysGluLeuPheAspGlyGluProValLeuAsnPheAlaIleValAspLys 272  
 Db 675 CAGTCTGTGGCCCTGCCCATG---TGAAGATGATGCTCAACATGATGTCAGCCACT 731  
 Qy 273 LeuPheTyrAsnAlaProLysMetSerLeuLeuAspTyrLeuLeuIleValAspPro 292  
 Db 732 GCCTTTTAAAGCA-----CAGCAGTGTAGTTCATGTGTGAGGTGTGGACATC 795  
 Qy 293 GlnSerCysAsnAspAspValArgLysAspLeuLysThrLysLeuMetAlaGlyLysMet 312  
 Db 786 AGGACATAGATGACAGCCCAAG----- 809  
 Qy 313 ThrIleArgGlnAlaAlaArgProArgIleArgGlnLeuLeuGluAsnLeuLysLeuLys 332  
 Db 810 CCCCTCAGGACTCTCAGCGCGTTCGCTTACCAGGAGATCAAGGCGCTGAAGGTGAA 869  
 Qy 333 -----CysAlaGluValTrpAspAsn-----GluMetSerArgLeuThr 345  
 Db 870 GTCACCCACTGTGGACAGATGAAGAGAGTAGTACCGGTGTGTAAATTACCGTCCGCCCT 929  
 Qy 346 GluArgHisLeuThrPheLeuAspLeuCysGluGluAsnSerLeuValTyrLysValThr 365

Db 930 GCTAGCCATCAGACATTCCCTTACAGCTGGAGAGTGACAGACTGTG----- 977  
Qy 366 GlyLysSerAspArgGlyArgAsnAlaLysLysTyrAspThrLeuPheLysLeuTyr 385  
Db 978 -----GAGTGCACAGTGGCAGCTATTTC 1001  
Qy 386 GluGluAsnLysLysPhe-----IleGluPheProHisLeuProLeuValLysValLysSer 404  
Db 1002 AAGCAGAAATATAACCTTCACTCAAGTATCCCATCTGCTGCTCCCTACAAAGTTGGCCAG 1061  
Qy 405 GlyAlaLysGluTyrAlaValProMetGluHisLeuGluValHisGluLysProGluArg 424  
Db 1062 GAACAAAGCATCTACCTTCCCTAGAGCTGTGTAACATT---GTGGCTGGCAGCGC 1118  
Qy 425 TyrLysAsnArgIleAspLeuValMetGlnAspLysPheLysLysArgAlaThrArgLys 444  
Db 1119 TGTATTAAAGAGTCAGCGCAGCAACACAGACCTCGACCATGATGATAAGGCCACAGCTAGATCC 1178  
Qy 445 ProHisAspTyrLysGluAsnThrLeuLysMetLeuLysGluLeuAspPheSerSerGlu 464  
Db 1179 GCTCCACAGACAGAGGAGAGATCAGTCGCTGATGAAGATGCGACCTCAACTTAGAT 1238  
Qy 465 GluLeuAsnPheValGluArgPheGlyLeuCysSerLysLeuGlnMetIleGluCysPro 484  
Db 1239 CCC-----TACATCCAGGAATTGGGATCAAGTGAAGATGACATGACGAGGTGACA 1292  
Qy 485 GlyLysValLeuLysGluProMetLeuValAsnSerValAsnGluGlnIleLysMetThr 504  
Db 1293 GGGCGAGTGTGCGCGCGCCCATCTTGCACTAGTCAGCGCGCGCAACCGGCATTGCCACA 1352  
Qy 505 Pro-----ValIleArgGlyPheGlnGluLysGlnLeu---AsnValVal 518  
Db 1353 CCCAATCAGGGTGTCTGGACATGCGGGGG-----AAACAGTTCTCAATGGGATT 1403  
Qy 519 ProGluLysGluLeuCysCysAlaValPheValValAsnGluThrAlaGlyAsnProCys 538  
Db 1404 GAGATCAAGTCTGGCCATCGCTGCTTCGCCACCCCAAAA-----CAGTGT 1451  
Qy 539 LeuGluGluAsnAspValValLysPheTyrThrGluLeuLeuGly----- 553  
Db 1452 CGAGAGAG-----GTGCTCAAGAACTTCACAGACCGCTGCGGAGATTTCCAGGAT 1505  
Qy 554 ----GlyCysLysPheArgGlyIleArgIleGlyAlaAsnGluAsnArgGlyAlaGlnSe 572  
Db 1506 GCGGGAGTGCCTATCCAGGGTCAACCTTGTCTGCAAAATATGACAGGGGGCA-GACAG 1564  
Qy 572 rIleMetTyrAspAlaThrLysAsnGluTyrAlaPheTyrLysAsnCysThrLeuAsnTh 592  
Db 1565 CGTG----- 1568  
Qy 592 rGlyIleGlyArgPheGluIleAlaAlaThrGluAlaLysAsnMetPheGluArgLeuPr 612  
Db 1569 -----GAGCCTATGTTCCGGCATCTCAA 1591  
Qy 612 oAspLysGluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAsnAlaTy 632  
Db 1592 GAACACCTACTAGGCTGCGAGTCACTATTGTCATCTCCCGAGGAAGACGCGGTGTA 1651  
Qy 632 rGlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAsnGlnHisThrSe 652  
Db 1652 TGCTGAGGTGAACGTCGAGATACACTCTTGGGATGGCTACCGAGTGTGTGAGGT 1711  
Qy 652 rGluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePh 672  
Db 1712 GAAGAACGTTGGTCAAGACC-----TCACCTCAGACTCT 1744  
Qy 672 eTyrGlnIleAlaLeuLysIleAsnAlaLysLeuGlyGlyIleAsnGlnGluLeuAspTr 692  
Db 1745 GTCCAACTCTGCTCAAGATCAATGTCAACTTGGTGGCATTAACAGCATCTCTA----- 1799  
Qy 692 pSerGluIleAlaGluIleSerProGluLysGluArgArgLysThrMetProLeuTh 712  
Db 1800 -----GTCCACACACGAGCTCTGCGGTTTTTCAACAGCAGCATG-- 1838

Qy 712 rMetTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIle 732  
Db 1839 -ATATTCTGGAGCAGATGTTACACACCCCGCAGAGGGATGGAAAAAACCTTCTAT 1897  
Qy 732 eAlaAlaValValAlaSerIleAsnProGlyGlyThrIleTyrArgAsnMetIleValTh 752  
Db 1898 CACAGCAGTGGTAGGAGTAGTATGGATGCCACCCCGAGCGCATCTGTGTACTGTGCGGT 1957  
Qy 752 rGlnGluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIle 772  
Db 1958 ACAG-----COACCACCGCAGAGAT 1978  
Qy 772 eLeuGlu-----AlaLysPheValLys---LeuLeuArgGluPheAlaGluAsnAsnAs 789  
Db 1979 CATTTGAAGACTTGTCTACATGTTGCTGAGCTCTCATCAATTCACAAAGTCCACCG 2038  
Qy 789 pAsnArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLe 809  
Db 2039 TTTCAAG---CCTACCCGCATCATCTTCTCCGAGATGGGGTGCCTGAAGCCAGCTACC 2095  
Qy 809 uArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGl 829  
Db 2096 CCAGATACTCCACTATGAGCTACTGGCCATTGCTGATGCTGCTGATCAAACTG----- 2147  
Qy 829 uArgAspGlyGluAspProGluProLysTyrThrPheIleValIleGlnLysArgHisAs 849  
Db 2148 -----GAAAGAGACTTACCAGCTGGATCACTTATATTGTGTGTCAGAAACGCCATCA 2200  
Qy 849 nThrArgLeuLeuArgArgMetGluLysAspLysProValValAsnLysAspLeuThrPr 869  
Db 2201 CACCCCGCTTTTC---TGCTGACAAAGATGGCGAAATGGGAAGAGTGGTAACATCCC 2257  
Qy 869 oAlaGluThrAspValAlaValAlaValLysGlnTrpGluGluAspMetLysGluSe 889  
Db 2258 AGCT----- 2261  
Qy 889 rLysGluThrGlyIleValAsnProSerSerGlyThrThrValAspLysLeuIleValSe 909  
Db 2262 -----GGACCAAGTGGACCAACATCACCCA 2290  
Qy 909 rLysTyrLysPheAspPheLeuAlaSerHisHisGlyValLeuGlyThrSerArgPr 929  
Db 2291 CCCATTGATTGACTTCTATCTGTGCACCCAGCAGGATCCAGGCCACCGACCC 2350  
Qy 929 oGlyHisTyrThrValMetTyrAspLysGlyMetSerGlnAspGluValTyrLysMe 949  
Db 2351 ATCCCATTAATGTTCTTGGATGACAAACCGTTTCACAGCAGATGAGCTCCAGATCCT 2410  
Qy 949 tThrTyrGlyLeuAlaPheLeuSerAlaArgCysArgLysPheProIleSerLeuProValPr 969  
Db 2411 GAGCTACAGCTGTGCCACACTTACGTACGATGCACACGCTCTGTCTCTATCCAGCACC 2470  
Qy 969 oValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGl 989  
Db 2471 TGCTTACTATGCCCGCTGTGCTGCTTTCGGGCACGA---TACCACCTGTGTGACAAAGGA 2527  
Qy 989 uHisTyrIleGlyAsp 994  
Db 2528 GCATGACAGTGGAGAG 2543

## RESULT 15

AX877133  
LOCUS Sequence 12038 from Patent EP1074617.  
ACCESSION AX877133  
VERSION AX877133.1 GI:40031869  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.



```

Qy 539 LeuGluGluAenAspValValValPheTyrThrGluLeuIle-Gly----- 553
Db 1452 CGAGAAGAG-----GTCTCAAGAACTTCACAGCAGCTGCGGAAGATTTCACAGGAT 1505
Qy 554 ----GlyCysLysPheArgGlyIleArgIleGlyAlaAenGluAenArgGlyAlaGlnSe 572
Db 1506 GCGGGGATGCTATCCAGGGTCAACCTTGTTCTGCAAAATATGCACAGGGGGCA-GACAG 1564
Qy 572 rIleMetTyrAspAlaThrLysAenGluTyrAlaPheTyrLysAenCysThrLeuAenTh 592
Db 1565 CGTG----- 1568
Qy 592 rGlyIleGlyArgPheGluIleAlaAlaThrGluAlaLysAenMetPheGluArgLeuPr 612
Db 1569 -----GAGCTATGTTCGCGCATCTCAA 1591
Qy 612 oAspLysGluGlnLysValLeuMetPheIleIleSerLysArgGlnLeuAenAlaTy 632
Db 1592 GNACACCTACTCAGGGCTGCAGCTCATTTATGTCTATCTGCCAGGGAAGCGCGGTGA 1651
Qy 632 rGlyPheValLysHisTyrCysAspHisThrIleGlyValAlaAenGlnHisIleThrSe 652
Db 1652 TGCTGAGGTGAACGTGCGGAGATACACTCTTGGGAATGGCTACGCAGTGTGTGCAGGT 1711
Qy 652 rGluThrValThrLysAlaLeuAlaSerLeuArgHisGluLysGlySerLysArgIlePh 672
Db 1712 GAGAAGCTGTCTAAGACC-----TCACCTCAGACTCT 1744
Qy 672 eTyrGlnIleAlaLeuLysIleAenAlaLysLeuGlyGlyIleAenGlnGluLeuAspTr 692
Db 1745 GTCCAACTCTGCCTCAAGATCAATGTCAAATCTGGTGGCATTAACAGCATCTTA----- 1799
Qy 692 pSerGluIleAlaGluSerProGluLysGluArgLysThrMetProLeuTh 712
Db 1800 -----GTCCCAACACAGCGCTCTGCGGTTTTTCAACAGCCAGTG-- 1838
Qy 712 rMetTyrValGlyIleAspValThrHisProThrSerTyrSerGlyIleAspTyrSerIl 732
Db 1839 -ATATCTCGGAGCAGATGTACACACCCCGAGCGGGATGCGGAAAAAACCTCTCTAT 1897
Qy 732 eAlaAlaValAlaSerIleAenProGlyThrIleTyrArgAenMetIleValTh 752
Db 1898 CACAGCAGTGTAGGCAGTATGATGCCACCCAGCCGATACTGTGCTACTGTGCGGT 1957
Qy 752 rGlnGluGluCysArgProGlyGluArgAlaValAlaHisGlyArgGluArgThrAspIl 772
Db 1958 ACAG-----CGACCACGCAAGAGAT 1978
Qy 772 eLeuGlu-----AlaLysPheValLys---LeuLeuArgGluPheAlaGluAenAenAs 789
Db 1979 CATGGAAGACTTGTCTACATGGTGGTGGTGGCTCTCTCATCAATTTCTACAAGTCCACCCG 2038
Qy 789 pAenArgAlaProAlaHisIleValValTyrArgAspGlyValSerAspSerGluMetLe 809
Db 2039 TTTCAG--CCTACCGCATCATCTCTCGCGAGATGGGTGCTTGAAGGCCAGCTACC 2095
Qy 809 uArgValSerHisAspGluLeuArgSerLeuLysSerGluValLysGlnPheMetSerGl 829
Db 2096 CCAGATACTCCACTATGAGCTACTGGCCATTGCGTATGCGTGCATCAAACTG----- 2147
Qy 829 uArgAspGlyGluAspProGluProlLysTyrThrPheIleValIleGlnLysArgHisAs 849
Db 2148 -----GAAAAGACTTACCAGCTGGGATCATCTTATATTGTGTGCGAAGAACGCCATCA 2200
Qy 849 nThrArgLeuLeuArgArgMetGluLysAspLysProValValAenLysAspLeuThrPr 869
Db 2201 CACCGCCTTTC---TGTCTGACAGAAATGGCGAATTTGGGAAGAGTGGTAACATCCC 2257
Qy 869 oAlaGluThrAspValAlaValAlaLysGlnTrpGluGluAspMetLysGluSe 889
Db 2258 AGCT----- 2261
Qy 889 rLysGluThrGlyIleValAenProSerSerGlyThrThrValAspLysLeuIleValSe 909

```

---

```

Db 2262 -----GGGACCACAGTGGACACCAACATCACCCA 2290
Qy 909 rLysTyrLysPheAspPheLeuAlaSerHisGlyValLeuGlyThrSerArgPr 929
Db 2291 CCCATTTGAGTTTGACTTCTATCTGTGCACGCCAGGAGCATCCAGGCGACGCCGACC 2350
Qy 929 oGlyHisTyrThrValMetTyrAspAspLysGlyMetSerGlnAspGluValTyrLysMe 949
Db 2351 ATCCCATTTACTATCTTTTGGGATGACAAACCGTTTCACAGCAGATGAGCTCCAGATCCT 2410
Qy 949 tThrTyrGlyLeuAlaPheLeuSerAlaAtqCysArgLysProIleSerLeuProValPr 969
Db 2411 GAGGTACACAGTGTGCCACACTTACGTACGATGCACACGCTCTCTCTATCCAGCACC 2470
Qy 969 oValHisTyrAlaHisLeuSerCysGluLysAlaLysGluLeuTyrArgThrTyrLysGl 989
Db 2471 TGCTACTATGCCCGCTGTGGTTCGGGACGA---TACCACCTGGTGGACAGGA 2527
Qy 989 uHisTyrIleGlyAsp 994
Db 2528 GCATGACAGTGGAGAG 2543

```

Search completed: July 6, 2006, 02:08:28  
Job time : 12644 secs

**THIS PAGE BLANK (except)**